BUNDESREPUBLIK DEUTSCHLAND

Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen:	199 31 424.1
Anmeldetag:	08. Juli 1999
Anmelder/Inhaber:	BASF Aktiengesellschaft, Ludwigshafen/DE
Bezeichnung:	Corynebacterium glutamicum genes encoding novel proteins
IPC:	C 07 K 14/34

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 20. Juli 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

CERTIFIED COPY OF
PRIORITY DOCUMENT

Hiebinger

A 9161
06/00
EDV-L

BEST AVAILABLE COPY

14.07.99

2/106

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION

Abstract of the Disclosure

Isolated nucleic acid molecules, designated HA nucleic acid molecules, which encode novel HA proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing HA nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated HA proteins, mutated HA proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of HA genes in this organism.

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Summary of the Invention

This invention provides novel nucleic acid molecules which may be used to identify or classify *Corynebacterium glutamicum* or related species of bacteria. *C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The nucleic acid molecules therefore can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. While *C. glutamicum* itself is nonpathogenic, it is related to other Corynebacterium species, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria), which are important human pathogens. The ability to identify the presence of Corynebacterium species therefore also can have significant clinical relevance, e.g., diagnostic applications. Further, these nucleic acid molecules may serve as reference points for the mapping of the *C. glutamicum* genome, or of genomes of related organisms.

These novel nucleic acid molecules encode proteins, referred to herein as homeostasis and adaptation (HA) proteins. These HA proteins are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or in the ability of this microorganism to adapt to different environmental conditions. Given the availability of cloning vectors for use in *Corynebacterium*

glutamicum, such as those disclosed in Sinskey et al., U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al, *J. Bacteriol* 162: 591-597 (1985); Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984); and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (e.g. phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (e.g., amino acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various *in vitro* industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C. glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (e.g., regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C. glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as HA proteins, which are capable of, for example, performing a function involved in the maintenance of homeostasis in *C. glutamicum*, or of

participating in the ability of this microorganism to adapt to different environmental conditions. Nucleic acid molecules encoding an HA protein are referred to herein as HA nucleic acid molecules. In a preferred embodiment, an HA protein participates in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or possesses a *C. glutamicum* enzymatic or proteolytic activity. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs) comprising a nucleotide sequence encoding an HA protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of HA-encoding nucleic acids (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g., sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an HA activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire

amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C glutamicum* and encodes a protein (e.g., an HA fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* HA protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an HA protein by culturing the host cell in a suitable medium. The HA protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an HA gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated HA sequence as a transgene. In another embodiment, an endogenous HA gene within the genome of the microorganism has been altered, e.g., functionally disrupted, by homologous recombination with an altered HA gene. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

Still another aspect of the invention pertains to an isolated HA protein or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated HA protein or portion thereof can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this

microorganism to different environmental conditions. In another preferred embodiment, the isolated HA protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions.

The invention also provides an isolated preparation of an HA protein. In preferred embodiments, the HA protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%. and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated HA protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated HA protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98,%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of HA proteins also have one or more of the HA bioactivities described herein.

The HA polypeptide, or a biologically active portion thereof, can be operatively linked to a non-HA polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the HA protein alone. In other preferred embodiments, this fusion protein participates in the maintenance of homeostasis in *C. glutamicum*, or performs a function involved in the adaptation of this microorganism to different environmental conditions. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an HA nucleic acid molecule of the invention, such that a fine chemical is

produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an HA nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates HA protein activity or HA nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C glutamicum* processes involved in cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activities. The agent which modulates HA protein activity can be an agent which stimulates HA protein activity or HA nucleic acid expression. Examples of agents which stimulate HA protein activity or HA nucleic acid expression include small molecules, active HA proteins, and nucleic acids encoding HA proteins that have been introduced into the cell. Examples of agents which inhibit HA activity or expression include small molecules and antisense HA nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant HA gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

Detailed Description of the Invention

The present invention provides HA nucleic acid and protein molecules which are involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C glutamicum* enzymatic or proteolytic activity. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (e.g., where overexpression or

optimization of activity of a protein involved in the production of a fine chemical (e.g., an enzyme) has a direct impact on the yield, production, and/or efficiency of production of a fine chemical from the modified *C. glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (e.g., where modulation of the activity or number of copies of a *C. glutamicum* aromatic or aliphatic modification or degradation protein results in an increase in the viability of *C glutamicum* cells, which in turn permits increased production in a large-scale culture setting). Aspects of the invention are further explicated below.

I. Fine Chemicals

The term "fine chemical" is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g., hyaluronic acid and trehalose), aromatic compounds (e.g., aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in

proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3[rd] edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosynthesis, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino aids – technical production and use, p. 466-502 in Rehm et al. (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E.(1978) *Ann Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α-ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and

resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β-carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications

of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B_1) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B_2) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B_6' (e.g., pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B_5), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B_{12}) and porphyrines belong to a group of chemicals characterized by a tetrapyrole ring system The biosynthesis of vitamin B_{12} is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B_6, pantothenate, and biotin. Only Vitamin B_{12} is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA

synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (i.e., AMP) or as coenzymes (i.e., FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (e.g. Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med Res Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol* 5: 752-757; (1995) *Biochem Soc Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (e.g., thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (e.g., ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (e.g., IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and Related Compounds in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction

reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in α, α-1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto et al., (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2. 293-314; and Shiosaka, M. (1997) J. Japan 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

<u>*II.* Maintenance of Homeostasis in *C. glutamicum* and Environmental Adaptation</u>

The metabolic and other biochemical processes by which cells function are sensitive to environmental conditions such as temperature, pressure, solute concentration, and availability of oxygen. When one or more such environmental condition is perturbed or altered in a fashion that is incompatible with the normal functioning of these cellular processes, the cell must act to maintain an intracellular environment which will permit them to occur despite the hostile extracellular environment. Gram positive bacterial cells, such as *C. glutamicum* cells, have a number of mechanisms by which internal homeostasis may be maintained despite unfavorable extracellular conditions. These include a cell wall, proteins which are able to degrade possibly toxic aromatic and aliphatic compounds, mechanisms of proteolysis whereby misfolded or misregulated proteins may be rapidly destroyed, and catalysts which permit intracellular reactions to occur which would not normally take place under the conditions optimal for bacterial growth.

Aside from merely surviving in a hostile environment, bacterial cells (e.g. *C glutamicum* cells) are also frequently able to adapt such that they are able to take advantage of such conditions. For example, cells in an environment lacking desired carbon sources may be able to adapt to growth on a less-suitable carbon source Also, cells may be able to utilize less desirable inorganic compounds when the commonly utilized ones are unavailable. *C glutamicum* cells possess a number of genes which permit them to adapt to utilize inorganic and organic molecules which they would normally not encounter under optimal growth conditions as nutrients and precursors for

metabolism. Aspects of cellular processes involved in homeostasis and adaptation are further explicated below.

A. Modification and Degradation of Aromatic and Aliphatic Compounds

Bacterial cells are routinely exposed to a variety of aromatic and aliphatic compounds in nature. Aromatic compounds are organic molecules having a cyclic ring structure, while aliphatic compounds are organic molecules having open chain structures rather than ring structures. Such compounds may arise as by-products of industrial processes (e.g., benzene or toluene), but may also be produced by certain microorganisms (e.g., alcohols). Many of these compounds are toxic to cells, particularly the aromatic compounds, which are highly reactive due to the high-energy ring structure. Thus, certain bacteria have developed mechanisms by which they are able to modify or degrade these compounds such that they are no longer hazardous to the cell. Cells may possess enzymes that are able to, for example, hydroxylate, isomerize, or methylate aromatic or aliphatic compounds such that they are either rendered less toxic, or such that the modified form is able to be processed by standard cellular waste and degradation pathways. Also, cells may possess enzymes which are able to specifically degrade one or more such potentially hazardous substance, thereby protecting the cell. Principles and examples of these types of modification and degradation processes in bacteria are described in several publications, e.g., Sahm, H. (1999) "Procaryotes in Industrial Production" in Lengeler, J.W. et al., eds. Biology of the Procaryotes, Thieme Verlag: Stuttgart; and Schlegel, H.G. (1992) Allgemeine Mikrobiologie, Thieme: Stuttgart).

Aside from simply inactivating hazardous aromatic or aliphatic compounds, many bacteria have evolved to be able to utilize these compounds as carbon sources for continued metabolism when the preferred carbon sources of the cell are not available. For example, Pseudomonas strains able to utilize toluene, benzene, and 1,10-dichlorodecane as carbon sources are known (Chang, B.V. et al. (1997) *Chemosphere* 35(12): 2807-2815; Wischnak, C. et al. (1998) *Appl Environ. Microbiol.* 64(9): 3507-3511; Churchill, S.A. et al. (1999) *Appl Environ. Microbiol.* 65(2): 549-552). There are similar examples from many other bacterial species which are known in the art.

The ability of certain bacteria to modify or degrade aromatic and aliphatic compounds has begun to be exploited. Petroleum is a complex mixture of chemicals which includes aliphatic molecules and aromatic compounds. By applying bacteria having the ability to degrade or modify these toxic compounds to an oil spill, for example, it is possible to eliminate much of the environmental damage with high efficiency and low cost (see, for example, Smith, M.R. (1990) "The biodegradation of aromatic hydrocarbons by bacteria"

Biodegradation 1(2-3): 191-206; and Suyama, T. et al. (1998) "Bacterial isolates degrading aliphatic polycarbonates," *FEMS Microbiol Lett.* 161(2): 255-261).

B Metabolism of Inorganic Compounds

Cells (e.g., bacterial cells) contain large quantities of different molecules, such as water, inorganic ions, and organic substances (e.g., proteins, sugars, and other macromolecules). The bulk of the mass of a typical cell consists of only 4 types of atoms: carbon, oxygen, hydrogen, and nitrogen. Although they represent a smaller percentage of the content of a cell, inorganic substances are equally as important to the proper functioning of the cell. Such molecules include phosphorous, sulfur, calcium, magnesium, iron, zinc, manganese, copper, molybdenum, tungsten, and cobalt. Many of these compounds are critical for the construction of important molecules, such as nucleotides (phosphorous) and amino acids (nitrogen and sulfur). Others of these inorganic ions serve as cofactors for enzymic reactions or contribute to osmotic pressure. All such molecules must be taken up by the bacterium from the surrounding environment.

For each of these inorganic compounds it is desirable for the bacterium to take up the form which can be most readily used by the standard metabolic machinery of the cell. However, the bacterium may encounter environments in which these preferred forms are not readily available. In order to survive under these circumstances, it is important for bacteria to have additional biochemical mechanisms which are able to convert less metabolically active but readily available forms of these inorganic compounds to ones which may be used in cellular metabolism. Bacteria frequently possess a number of genes encoding enzymes for this purpose, which are not expressed unless the desired inorganic species are not available. Thus, these genes for the metabolism of various inorganic compounds serve as another tool which bacteria may use to adapt to suboptimal environmental conditions.

After carbon, the most important element in the cell is nitrogen. A typical bacterial cell contains between 12-15% nitrogen. It is a constituent of amino acids and nucleotides, as well as many other important molecules in the cell. Further, nitrogen may serve as a substitute for oxygen as a terminal electron acceptor in energy metabolism. Good sources of nitrogen include many organic and inorganic compounds, such ammonia gas or ammonia salts (e.g., NH_4Cl, $(NH_4)_2SO_4$, or NH_4OH), nitrates, urea, amino acids, or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract, etc. Ammonia nitrogen is fixed by the action of particular enzymes: glutamate dehydrogenase, glutamine synthase, and glutamine-2-oxoglutarate aminotransferase. The transfer of amino-nitrogen from one organic molecule to another is accomplished by the aminotransferases, a class of enzymes which transfer one amino group from an alpha-amino acid to an alpha-keto acid. Nitrate may be reduced via nitrate reductase, nitrite reductase,

and further redox enzymes until it is converted to molecular nitrogen or ammonia, which may be readily utilized by the cell in standard metabolic pathways.

Phosphorous is typically found intracellularly in both organic and inorganic forms, and may be taken up by the cell in either of these forms as well, though most microorganisms preferentially take up inorganic phosphate. The conversion of organic phosphate to a form which the cell can utilize requires the action of phosphatases (e.g., phytases, which hydrolyze phyate-yielding phosphate and inositol derivatives). Phosphate is a key element in the synthesis of nucleic acids, and also has a significant role in cellular energy metabolism (e.g., in the synthesis of ATP, ADP, and AMP).

Sulfur is a requirement for the synthesis of amino acids (e.g., methionine and cysteine), vitamins (e.g., thiamine, biotin, and lipoic acid) and iron sulfur proteins. Bacteria obtain sulfur primarily from inorganic sulfate, though thiosulfate, sulfite, and sulfide are also commonly utilized. Under conditions where these compounds may not be readily available, many bacteria express genes which enable them to utilize sulfonate compounds such as 2-aminosulfonate (taurine) (Kertesz, M.A. (1993) "Proteins induced by sulfate limitation in *Escherichia coli*, *Pseudomonas putida*, or *Staphylococcus aureus*." *J. Bacteriol.* 175: 1187-1190).

Other inorganic atoms, e.g., metal or calcium ions, are also critical for the viability of cells. Iron, for example, plays a key role in redox reactions and is a cofactor of iron-sulfur proteins, heme proteins, and cytochromes. The uptake of iron into bacterial cells may be accomplished by the action of siderophores, chelating agents which bind extracellular iron ions and translocate them to the interior of the cell. For reference on the metabolism of iron and other inorganic compounds, see: Lengeler et al. (1999) Biology of Prokaryotes, Thieme Verlag: Stuttgart; Neidhardt, F.C. et al., eds. *Escherichia coli* and *Salmonella*. ASM Press: Washington, D.C.; Sonenshein, A.L. et al., eds. (199?) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM Press: Washington, D.C.; Voet, D. and Voet, J.G. (1992) Biochemie, VCH: Weinheim; Brock, T.D. and Madigan, M.T. (1991) Biology of Microorgansisms, 6[th] ed. Prentice Hall: Englewood Cliffs, p. 267-269; Rhodes, P.M. and Stanbury, P.F. Applied Microbial Physiology – A Practical Approach, Oxford Univ. Press: Oxford.

C. Enzymes and Proteolysis

The intracellular conditions for which bacteria such as *C glutamicum* are optimized are frequently not conditions under which many biochemical reactions would normally take place. In order to make such reactions proceed under physiological conditions, cells utilize enzymes. Enzymes are proteinaceous biological catalysts, spatially orienting reacting molecules or providing a specialized environment such that

the energy barrier to a biochemical reaction is lowered. Different enzymes catalyze different reactions, and each enzyme may be the subject of transcriptional, translational, or posttranslational regulation such that the reaction will only take place under appropriate conditions and at specified times. Enzymes may contribute to the degradation (e.g., the proteases), synthesis (e.g., the synthases), or modification (e.g., transferases or isomerases) of compounds, all of which enable the production of necessary compounds within the cell. This, in turn, contributes to the maintenance of cellular homeostasis.

However, the fact that enzymes are optimized for activity under the physiological conditions at which the bacterium is most viable means that when environmental conditions are perturbed, there is a significant possibility that enzyme activity will also be perturbed. For example, changes in temperature may result in aberrantly folded proteins, and the same is true for changes of pH – protein folding is largely dependent on electrostatic and hydrophobic interactions of amino acids within the polypeptide chain, so any alteration to the charges on individual amino acids (as might be brought about by a change in cellular pH) may have a profound effect on the ability of the protein to correctly fold. Changes in temperature effectively change the amount of kinetic energy that the polypeptide molecule possesses, which affects the ability of the polypeptide to settle into a correctly folded, energetically stable configuration. Misfolded proteins may be harmful to the cell for two reasons. First, the aberrantly folded protein may have a similarly aberrant activity, or no activity whatsoever. Second, misfolded proteins may lack the conformational regions necessary for proper regulation by other cellular systems and thus may continue to be active but in an uncontrolled fashion.

The cell has a mechanism by which misfolded enzymes and regulatory proteins may be rapidly destroyed before any damage occurs to the cell: proteolysis. Proteins such as those of the la/lon family and those of the Clp family specifically recognize and degrade misfolded proteins (see, e.g., Sherman, M.Y., Goldberg, A.L. (1999) *EXS* 77: 57-78 and references therein and Porankiewicz J. (1999) *Molec. Microbiol.* 32(3): 449-58, and references therein; Neidhardt, F.C., et al. (1996) *E. coli* and *Salmonella*, ASM Press: Washington, D.C. and references therein; and Pritchard, G.G., and Coolbear, T. (1993) *FEMS Microbiol.* Rev. 12(1-3): 179-206 and references therein). These enzymes bind to misfolded or unfolded proteins and degrade them in an ATP-dependent manner. Proteolysis thus serves as an important mechanism employed by the cell to prevent damage to normal cellular functions upon environmental changes, and it further permits cells to survive under conditions and in environments which would otherwise be toxic due to misregulated and/or aberrant enzyme or regulatory activity.

Proteolysis also has important functions in the cell under optimal environmental conditions. Within normal metabolic processes, proteases aid in the hydrolysis of peptide bonds, in the catabolism of complex molecules to provide necessary degradation products, and in protein modification. Secreted proteases play an important role in the catabolism of external nutrients even prior to the entry of these compounds into the cell. Further, proteolytic activity itself may serve regulatory functions; sporulation in *B. subtilis* and cell cycle progression in *Caulobacter* spp. are known to be regulated by key proteolytic events in each of these species (Gottesman, S. (1999) *Curr. Opin. Microbiol.* 2(2): 142-147). Thus, proteolytic processes are key for cellular survival under both suboptimal and optimal environmental conditions, and contribute to the overall maintenance of homeostasis in cells.

D. Cell Wall Production and Rearrangements

While the biochemical machinery of the cell may be able to readily adapt to different and possibly unfavorable environments, cells still require a general mechanism by which they may be protected from the environment. For many bacteria, the cell wall affords such protection, and also plays roles in adhesion, cell growth and division, and transport of desired solutes and waste materials.

In order to function, cells require intracellular concentrations of metabolites and other molecules that are substantially higher than those of the surrounding media. Since these metabolites are largely prevented from leaving the cell due to the presence of the hydrophobic membrane, the tendency of the system is for water molecules to enter the cell from the external medium such that the interior concentrations of solutes match the exterior concentrations. Water molecules are readily able to cross the cellular membrane, and this membrane is not able to withstand the resulting swelling and pressure, which may lead to osmotic lysis of the cell. The rigidity of the cell wall greatly improves the ability of the cell to tolerate these pressures, and offers a further barrier to the unwanted diffusion of these metabolites and desired solutes from the cell. Similarly, the cell wall also serves to prevent unwanted material from entering the cell.

The cell wall also participates in a number of other cellular processes, such as adhesion and cell growth and division. Due to the fact that the cell wall completely surrounds the cell, any interaction of the cell with its surroundings must be mediated by the cell wall. Thus, the cell wall must participate in any adherence of the cell to other cells and to desired surfaces. Further, the cell cannot grow or divide without concomitant changes in the cell wall. Since the protection that the wall affords requires its presence during growth, morphogenesis and multiplication, one of the key steps in cell division is cell wall synthesis within the cell such that a new cell divides from the old. Thus, frequently cell wall

biosynthesis is regulated in tandem with cell growth and cell division (see, e.g., Sonenshein, A.L. et al, eds. (1993) *Bacillus subtilis* and Other Gram-Positive Bacteria, ASM: Washington, D.C.).

The structure of the cell wall varies between gram-positive and gram-negative bacteria. However, in both types, the fundamental structural unit of the wall remains similar: an overlapping lattice of two polysaccharides, N-acetyl glucosamine (NAG) and N-acetyl muramic acid (NAM) which are cross-linked by amino acids (most commonly L-alanine, D-glutamate, diaminopimelic acid, and D-alanine), termed 'peptidoglycan'. The processes involved in the synthesis of the cell wall are known (see, e.g., Michal, G., ed. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York).

In gram-negative bacteria, the inner cellular membrane is coated by a single-layered peptidoglycan (approximately 10 nm thick), termed the murein-sacculus. This peptidoglycan structure is very rigid, and its structure determines the shape of the organism. The outer surface of the murein-sacculus is covered with an outer membrane, containing porins and other membrane proteins, phospholipids, and lipopolysaccharides. To maintain a tight association with the outer membrane, the gram-negative cell wall also has interspersed lipid molecules which serve to anchor it to the surrounding membrane.

In gram-positive bacteria, such as *Corynebacterium glutamicum*, the cytoplasmic membrane is covered by a multi-layered peptidoglycan, which ranges from 20-80 nm in thickness (see, e.g., Lengeler et al. (1999) Biology of Prokaryotes Thieme Verlag: Stuttgart, p. 913-918, p. 875-899, and p. 88-109 and references therein). The gram-positive cell wall also contains teichoic acid, a polymer of glycerol or ribitol linked through phosphate groups. Teichoic acid is also able to associate with amino acids, and forms covalent bonds with muramic acid. Also present in the cell wall may be lipoteichoic acids and teichuronic acids. If present, cellular surface structures such as flagella or capsules will be anchored in this layer as well.

III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as HA nucleic acid and protein molecules, which participate in the maintenance of homeostasis in *C. glutamicum*, or which perform a function involved in the adaptation of this microorganism to different environmental conditions. In one embodiment, the HA molecules participate in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds, in the modification or degradation of aromatic or aliphatic compounds, or have an enzymatic or proteolytic activity. In a preferred embodiment, the activity of the HA molecules of

the present invention with regard to *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the HA molecules of the invention are modulated in activity, such that the *C. glutamicum* cellular processes in which the HA molecules participate (e.g., *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or enzymatic or proteolytic activity) are also altered in activity, resulting either directly or indirectly in a modulation of the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

The language, "HA protein" or "HA polypeptide" includes proteins which participate in a number of cellular processes related to *C. glutamicum* homeostasis or the ability of *C. glutamicum* cells to adapt to unfavorable environmental conditions. For example, an HA protein may be involved in *C. glutamicum* cell wall biosynthesis or rearrangements, in the metabolism of inorganic compounds in *C. glutamicum*, in the modification or degradation of aromatic or aliphatic compounds in *C. glutamicum*, or have a *C. glutamicum* enzymatic or proteolytic activity. Examples of HA proteins include those encoded by the HA genes set forth in Table 1 and Appendix A. The terms "HA gene" or "HA nucleic acid sequence" include nucleic acid sequences encoding an HA protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of HA genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an

organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound. The term "homeostasis" is art-recognized and includes all of the mechanisms utilized by a cell to maintain a constant intracellular environment despite the prevailing extracellular environmental conditions. A non-limiting example of such processes is the utilization of a cell wall to prevent osmotic lysis due to high intracellular solute concentrations. The term "adaptation" or "adaptation to an environmental condition" is art-recognized and includes mechanisms utilized by the cell to render the cell able to survive under nonpreferred environmental conditions (generally speaking, those environmental conditions in which one or more favored nutrients are absent, or in which an environmental condition such as temperature, pH, osmolarity, oxygen percentage and the like fall outside of the optimal survival range of the cell). Many cells, including *C. glutamicum* cells, possess genes encoding proteins which are expressed under such environmental conditions and which permit continued growth in such suboptimal conditions.

In another embodiment, the HA molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an HA protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by engineering enzymes which modify or degrade aromatic or aliphatic compounds such that these enzymes are increased or decreased in activity or number, it may be possible to modulate the production of one or more fine chemicals which are the modification or degradation products of these compounds. Similarly, enzymes involved in the metabolism of inorganic compounds provide key molecules (e.g. phosphorous, sulfur, and nitrogen molecules) for the biosynthesis of such fine chemicals as amino acids, vitamins, and nucleic acids. By altering the activity or number of these enzymes in *C. glutamicum*, it may be possible to increase the conversion of these inorganic compounds (or to use alternate inorganic compounds) to thus permit improved rates of incorporation of inorganic atoms into these fine chemicals. Genetic engineering of *C. glutamicum* enzymes involved in general cellular processes may also directly improve fine chemical production, since many of these enzymes directly modify fine chemicals (e.g., amino

acids) or the enzymes which are involved in fine chemical synthesis or secretion. Modulation of the activity or number of cellular proteases may also have a direct effect on fine chemical production, since many proteases may degrade fine chemicals or enzymes involved in fine chemical production or breakdown.

Further, the aforementioned enzymes which participate in aromatic/aliphatic compound modification or degradation, general biocatalysis, inorganic compound metabolism or proteolysis are each themselves fine chemicals, desirable for their activity in various *in vitro* industrial applications. By altering the number of copies of the gene for one or more of these enzymes in *C. glutamicum* it may be possible to increase the number of these proteins produced by the cell, thereby increasing the potential yield or efficiency of production of these proteins from large-scale *C. glutamicum* or related bacterial cultures.

The alteration of an HA protein of the invention may also indirectly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, by modulating the activity and/or number of those proteins involved in the construction or rearrangement of the cell wall, it may be possible to modify the structure of the cell wall itself such that the cell is able to better withstand the mechanical and other stresses present during large-scale fermentative culture. Also, large-scale growth of *C glutamicum* requires significant cell wall production. Modulation of the activity or number of cell wall biosynthetic or degradative enzymes may allow more rapid rates of cell wall biosynthesis, which in turn may permit increased growth rates of this microorganism in culture and thereby increase the number of cells producing the desired fine chemical.

By modifying the HA enzymes of the invention, one may also indirectly impact the yield, production, or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, many of the general enzymes in *C. glutamicum* may have a significant impact on global cellular processes (e.g., regulatory processes) which in turn have a significant effect on fine chemical metabolism. Similarly, proteases, enzymes which modify or degrade possibly toxic aromatic or aliphatic compounds, and enzymes which promote the metabolism of inorganic compounds all serve to increase the viability of *C. glutamicum*. The proteases aid in the selective removal of misfolded or misregulated proteins, such as those that might occur under the relatively stressful environmental conditions encountered during large-scale fermentor culture. By altering these proteins, it may be possible to further enhance this activity and to improve the viability of *C glutamicum* in culture. The aromatic/aliphatic modification or degradation proteins not only serve to detoxify these waste compounds (which may be

encountered as impurities in culture medium or as waste products from cells themselves), but also to permit the cells to utilize alternate carbon sources if the optimal carbon source is limiting in the culture. By increasing their number and/or activity, the survival of *C. glutamicum* cells in culture may be enhanced. The inorganic metabolism proteins of the invention supply the cell with inorganic molecules required for all protein and nucleotide (among others) synthesis, and thus are critical for the overall viability of the cell. An increase in the number of viable cells producing one or more desired fine chemicals in large-scale culture should result in a concomitant increase in the yield, production, and/or efficiency of production of the fine chemical in the culture.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* HA cDNAs and the predicted amino acid sequences of the *C glutamicum* HA proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins that participate in *C glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The HA protein or a biologically active portion or fragment thereof of the invention can participate in the maintenance of homeostasis in *C. glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections.

A Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode HA polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of HA-encoding nucleic acid (e.g., HA DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated HA nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* HA cDNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of

Appendix A). For example, mRNA can be isolated from normal endothelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an HA nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* HA cDNAs of the invention. This cDNA comprises sequences encoding HA proteins (i.e., "the coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA number having the designation "RXA" followed by 5 digits (i.e., RXA00009). Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequence in Appendix B designated RXA00009 is a translation of the coding region of the nucleotide sequence of nucleic acid molecule RXA00009 in Appendix A.

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., et al. (1998) *J. Bacteriol.* 180(12): 3159-

3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an HA protein. The nucleotide sequences determined from the cloning of the HA genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning HA homologues in other cell types and organisms, as well as HA homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone HA homologues. Probes based on the HA nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-

factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an HA protein, such as by measuring a level of an HA-encoding nucleic acid in a sample of cells, e.g., detecting HA mRNA levels or determining whether a genomic HA gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. Proteins involved in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or aliphatic compounds, or that have a *C. glutamicum* enzymatic or proteolytic activity, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an HA protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of HA protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

Portions of proteins encoded by the HA nucleic acid molecules of the invention are preferably biologically active portions of one of the HA proteins. As used herein, the term "biologically active portion of an HA protein" is intended to include a portion, e.g., a domain/motif, of an HA protein that can participate in the maintenance of homeostasis in *C. glutamicum*, or that can perform a function involved in the adaptation of this microorganism to different environmental conditions, or has an activity as set forth in Table 1. To determine whether an HA protein or a biologically active portion thereof can participate in *C. glutamicum* cell wall biosynthesis or rearrangements, metabolism of inorganic compounds, modification or degradation of aromatic or

aliphatic compounds, or has a *C. glutamicum* enzymatic or proteolytic activity, an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an HA protein can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the HA protein or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the HA protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same HA protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In addition to the *C. glutamicum* HA nucleotide sequences shown in Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of HA proteins may exist within a population (e.g., the *C. glutamicum* population). Such genetic polymorphism in the HA gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an HA protein, preferably a *C. glutamicum* HA protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the HA gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in HA that are the result of natural variation and that do not alter the functional activity of HA proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* HA cDNA of the invention can be isolated based on their homology to the *C. glutamicum* HA nucleic acid disclosed herein using the *C. glutamicum* cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the

nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C glutamicum* HA protein.

In addition to naturally-occurring variants of the HA sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded HA protein, without altering the functional ability of the HA protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the HA proteins (Appendix B) without altering the activity of said HA protein, whereas an "essential" amino acid residue is required for HA protein activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having HA activity) may not be essential for activity and thus are likely to be amenable to alteration without altering HA activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding HA proteins that contain changes in amino acid residues that are not essential for HA activity. Such HA proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the HA activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of

participating in the maintenance of homeostasis in *C. glutamicum*, or of performing a function involved in the adaptation of this microorganism to different environmental conditions, or has one or more of the activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an HA protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g.,

threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an HA protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an HA coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an HA activity described herein to identify mutants that retain HA activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding HA proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire HA coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an HA protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID RXA00009 comprises nucleotides 1 to 900). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding HA. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding HA disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of HA mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of HA mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of HA mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can

be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an HA protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms

specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave HA mRNA transcripts to thereby inhibit translation of HA mRNA. A ribozyme having specificity for an HA-encoding nucleic acid can be designed based upon the nucleotide sequence of an HA DNA molecule disclosed herein (i.e., RXA00009 in Appendix A). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an HA-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742. Alternatively, HA mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, HA gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an HA nucleotide sequence (e.g., an HA promoter and/or enhancers) to form triple helical structures that prevent transcription of an HA gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann N.Y. Acad Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an HA protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other

vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which are operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., HA proteins, mutant forms of HA proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of HA proteins in prokaryotic or eukaryotic cells. For example, HA genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. et al. (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. et al. (1991) "Heterologous gene expression in filamentous fungi" in: More

Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* –mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the HA protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant HA protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host

RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the HA protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

Alternatively, the HA proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the HA proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20. 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl. Acid. Res.* 12: 8711-8721.

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning. A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to HA mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid

or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an HA protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an HA protein or can be

introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an HA gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the HA gene. Preferably, this HA gene is a *Corynebacterium glutamicum* HA gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous HA gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous HA gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous HA protein). In the homologous recombination vector, the altered portion of the HA gene is flanked at its 5' and 3' ends by additional nucleic acid of the HA gene to allow for homologous recombination to occur between the exogenous HA gene carried by the vector and an endogenous HA gene in a microorganism. The additional flanking HA nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (e.g., by electroporation) and cells in which the introduced HA gene has homologously recombined with the endogenous HA gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an HA gene on a vector placing it under control of the lac operon permits expression of the HA gene only in the presence of IPTG. Such regulatory systems are well known in the art.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an HA protein. Accordingly, the invention further provides methods for producing HA proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an HA protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered HA protein) in a suitable medium until HA protein is produced. In another

embodiment, the method further comprises isolating HA proteins from the medium or the host cell.

C. Isolated HA Proteins

Another aspect of the invention pertains to isolated HA proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of HA protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of HA protein having less than about 30% (by dry weight) of non-HA protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-HA protein, still more preferably less than about 10% of non-HA protein, and most preferably less than about 5% non-HA protein. When the HA protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of HA protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of HA protein having less than about 30% (by dry weight) of chemical precursors or non-HA chemicals, more preferably less than about 20% chemical precursors or non-HA chemicals, still more preferably less than about 10% chemical precursors or non-HA chemicals, and most preferably less than about 5% chemical precursors or non-HA chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the HA protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* HA protein in a microorganism such as *C. glutamicum*.

An isolated HA protein or a portion thereof of the invention can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises

an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the maintenance of homeostasis in *C. glutamicum*, or to perform a function involved in the adaptation of this microorganism to different environmental conditions. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an HA protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the HA protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences of Appendix B. The preferred HA proteins of the present invention also preferably possess at least one of the HA activities described herein. For example, a preferred HA protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the maintenance of homeostasis in *C glutamicum*, or can perform a function involved in the adaptation of this microorganism to different environmental conditions, or which has one or more of the activities set forth in Table 1.

In other embodiments, the HA protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the HA protein is a protein which comprises an amino acid sequence which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the HA activities described herein. In another embodiment, the invention pertains to a full length *C glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an HA protein include peptides comprising amino acid sequences derived from the amino acid sequence of an HA protein, e.g., the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an HA protein, which include fewer amino acids than a full length HA protein or the full length protein which is homologous to an HA protein, and exhibit at

least one activity of an HA protein. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an HA protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an HA protein include one or more selected domains/motifs or portions thereof having biological activity.

SES proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the HA protein is expressed in the host cell. The HA protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an HA protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native HA protein can be isolated from cells (e.g., endothelial cells), for example using an anti-HA antibody, which can be produced by standard techniques utilizing an HA protein or fragment thereof of this invention.

The invention also provides HA chimeric or fusion proteins. As used herein, an HA "chimeric protein" or "fusion protein" comprises an HA polypeptide operatively linked to a non-HA polypeptide. An "HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an HA protein, whereas a "non-HA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the HA protein, e.g., a protein which is different from the HA protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the HA polypeptide and the non-HA polypeptide are fused in-frame to each other. The non-HA polypeptide can be fused to the N-terminus or C-terminus of the HA polypeptide. For example, in one embodiment the fusion protein is a GST-HA fusion protein in which the HA sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant HA proteins. In another embodiment, the fusion protein is an HA protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of an HA protein can be increased through use of a heterologous signal sequence.

Preferably, an HA chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with

conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An HA-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the HA protein.

Homologues of the HA protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the HA protein. As used herein, the term "homologue" refers to a variant form of the HA protein which acts as an agonist or antagonist of the activity of the HA protein. An agonist of the HA protein can retain substantially the same, or a subset, of the biological activities of the HA protein. An antagonist of the HA protein can inhibit one or more of the activities of the naturally occurring form of the HA protein, by, for example, competitively binding to a downstream or upstream member of a biochemical cascade which includes the HA protein, by binding to a target molecule with which the HA protein interacts, such that no functional interaction is possible, or by binding directly to the HA protein and inhibiting its normal activity.

In an alternative embodiment, homologues of the HA protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the HA protein for HA protein agonist or antagonist activity. In one embodiment, a variegated library of HA variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of HA variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HA sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of HA sequences therein. There are a variety of methods which can be used to produce libraries of potential HA homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding

the desired set of potential HA sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the HA protein coding can be used to generate a variegated population of HA fragments for screening and subsequent selection of homologues of an HA protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an HA coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the HA protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HA homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify HA homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated HA library, using methods well known in the art.

D Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of HA protein

regions required for function; modulation of an HA protein activity; modulation of the metabolism of one or more inorganic compounds; modulation of the modification or degradation of one or more aromatic or aliphatic compounds; modulation of cell wall synthesis or rearrangements; modulation of enzyme activity or proteolysis; and modulation of cellular production of a desired compound, such as a fine chemical.

The HA nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. Detection of such organisms is of significant clinical relevance.

Further, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The HA nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The processes involved in adaptation and the maintenance of homeostasis in which the molecules of the invention participate are utilized by a wide variety of species; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are

conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the HA nucleic acid molecules of the invention may result in the production of HA proteins having functional differences from the wild-type HA proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

The modulation of activity or number of HA proteins involved in cell wall biosynthesis or rearrangements may both impact the production, yield, and/or efficiency of production of one or more fine chemicals from *C glutamicum* cells. For example, by altering the activity of these proteins, it may be possible to modulate the structure or thickness of the cell wall. The cell wall serves in large measure as a protective device against osmotic lysis and external sources of injury; by modifying the cell wall it may be possible to increase the ability of *C. glutamicum* to withstand the mechanical and shear force stresses encountered by this microorganism during large-scale fermentor culture. Further, each *C. glutamicum* cell is surrounded by a thick cell wall, and thus, a significant portion of the biomass present in large scale culture consists of cell wall. By increasing the rate at which the cell wall is synthesized or by activating cell wall synthesis (through genetic engineering of the HA cell wall proteins of the invention) it may be possible to improve the growth rate of the microorganism. Similarly, by decreasing the activity or number of proteins involved in the degradation of cell wall or by decreasing the repression of cell wall biosynthesis, an overall increase in cell wall production may be achieved. An increase in the number of viable *C. glutamicum* cells (as may be accomplished by any of the foregoing described protein alterations) should result in increased numbers of cells producing the desired fine chemical in large-scale fermentor culture, which should permit increased yields or efficiency of production of these compounds from the culture.

The modulation of activity or number of C. glutamicum HA proteins that participate in the modification or degradation of aromatic or aliphatic compounds may also have direct or indirect impacts on the production of one or more fine chemicals from these cells. Certain aromatic or aliphatic modification or degradation products are desirable fine chemicals (e.g., organic acids or modified aromatic and aliphatic compounds); thus, by modifying the enzymes which perform these modifications (e.g., hydroxylation, methylation, or isomerization) or degradation reactions, it may be possible to increase the yields of these desired compounds. Similarly, by decreasing the activity or number of proteins involved in pathways which further degrade the modified

or breakdown products of the aforementioned reactions it may be possible to improve the yields of these fine chemicals from *C. glutamicum* cells in culture.

These aromatic and aliphatic modification and degradative enzymes are themselves fine chemicals. In purified form, these enzymes may be used to degrade aromatic and aliphatic compounds (e.g., toxic chemicals such as petroleum products), either for the bioremediation of polluted sites, for the engineered decomposition of wastes, or for the large-scale and economically feasible production of desired modified aromatic or aliphatic compounds or their breakdown products, some of which may be conveniently used as carbon or energy sources for other fine chemical-producing compounds in culture (see, e.g., Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin and references therein; and Roberts, S.M., ed. (1992-1996) Preparative Biotransformations, Wiley: Chichester, and references therein). By genetically altering these proteins such that their regulation by other cellular mechanisms is lessened or abolished, it may be possible to increase the overall number or activity of these proteins, thereby improving not only the yield of these fine chemicals but also the activity of these harvested proteins.

The modification of these aromatic and aliphatic modifying and degradation enzymes may also have an indirect effect on the production of one or more fine chemical. Many aromatic and aliphatic compounds (such as those that may be encountered as impurities in culture media or as waste products from cellular metabolism) are toxic to cells; by modifying and/or degrading these compounds such that they may be readily removed or destroyed, cellular viability should be increased. Further, these enzymes may modify or degrade these compounds in such a manner that the resulting products may enter the normal carbon metabolism pathways of the cell, thus rendering the cell able to use these compounds as alternate carbon or energy sources. In large-scale culture situations, when there may be limiting amounts of optimal carbon sources, these enzymes provide a method by which cells may continue to grow and divide using aromatic or aliphatic compounds as nutrients. In either case, the resulting increase in the number of *C. glutamicum* cells in the culture producing the desired fine chemical should in turn result in increased yields or efficiency of production of the fine chemical(s).

Modifications in activity or number of HA proteins involved in the metabolism of inorganic compounds may also directly or indirectly affect the production of one or more fine chemicals from *C. glutamicum* or related bacterial cultures. For example, many desirable fine chemicals, such as nucleic acids, amino acids, cofactors and vitamins (e.g., thiamine, biotin, and lipoic acid) cannot be synthesized without inorganic molecules such as phosphorous, nitrate, sulfate, and iron. The inorganic metabolism

proteins of the invention permit the cell to obtain these molecules from a variety of inorganic compounds and to divert them into various fine chemical biosynthetic pathways. Therefore, by increasing the activity or number of enzymes involved in the metabolism of these inorganic compounds, it may be possible to increase the supply of these possibly limiting inorganic molecules, thereby directly increasing the production or efficiency of production of various fine chemicals from *C. glutamicum* cells containing such altered proteins. Modification of the activity or number of inorganic metabolism enzymes of the invention may also render *C. glutamicum* able to better utilize limited inorganic compound supplies, or to utilize nonoptimal inorganic compounds to synthesize amino acids, vitamins, cofactors, or nucleic acids, all of which are necessary for continued growth and replication of the cell. By improving the viability of these cells in large-scale culture, the number of *C. glutomicum* cells producing one or more fine chemicals in the culture may also be increased, in turn increasing the yields or efficiency of production of one or more fine chemicals.

C. *glutamicum* enzymes for general processes are themselves desirable fine chemicals. The specific properties of enzymes (i.e., regio- and stereospecificity, among others) make them useful catalysts for chemical reactions *in vitro*. Either whole *C. glutamicum* cells may be incubated with an appropriate substrate such that the desired product is produced by enzymes in the cell, or the desired enzymes may be overproduced and purified from *C. glutamicum* cultures (or those of a related bacterium) and subsequently utilized in *in vitro* reactions in an industrial setting (either in solution or immobilized on a suitable immobile phase). In either situation, the enzyme can either be a natural *C glutamicum* protein, or it may be mutagenized to have an altered activity; typical industrial uses for such enzymes include as catalysts in the chemical industry (e.g., for synthetic organic chemistry) as food additives, as feed components, for fruit processing, for leather preparation, in detergents, in analysis and medicine, and in the textile industry (see, e.g.,Yamada, H. (1993) "Microbial reactions for the production of useful organic compounds," *Chimica* 47: 5-10; Roberts, S.M. (1998) Preparative biotransformations: the employment of enzymes and whole-cells in synthetic chemistry," *J. Chem. Soc. Perkin Trans*. 1: 157-169; Zaks, A. and Dodds, D.R. (1997) "Application of biocatalysis and biotransformations to the synthesis of pharmaceuticals," *DDT* 2: 513-531; Roberts, S.M. and Williamson, N.M. (1997) "The use of enzymes for the preparation of biologically active natural products and analogues in optically active form," *Curr. Organ Chemistry* 1: 1-20; Faber, K. (1995) Biotransformations in Organic Chemistry, Springer: Berlin; Roberts, S.M., ed. (1992-96) Preparative Biotransformations, Wiley: Chichester; Cheetham, P.S.J. (1995) "The applications of enzymes in industry" in : Handbook of Enzyme Biotechnology, 3rd ed.,

Wiseman, A., ed., Elis: Horwood, p. 419-552; and Ullmann's Encyclopedia of Industrial Chemistry (1987), vol. A9, Enzymes, p. 390-457). Thus, by increasing the activity or number of these enzymes, it may be possible to also increase the ability of the cell to convert supplied substrates to desired products, or to overproduce these enzymes for increased yields in large-scale culture. Further, by mutagenizing these proteins it may be possible to remove feedback inhibition or other repressive cellular regulatory controls such that greater numbers of these enzymes may be produced and activated by the cell, thereby leading to greater yields, production, or efficiency of production of these fine chemical proteins from large-scale cultures. Further, manipulation of these enzymes may alter the activity of one or more *C. glutamicum* metabolic pathways, such as those for the biosynthesis or secretion of one or more fine chemicals.

Mutagenesis of the proteolytic enzymes of the invention such that they are altered in activity or number may also directly or indirectly affect the yield, production, and/or efficiency of production of one or more fine chemicals from *C. glutamicum*. For example, by increasing the activity or number of these proteins, it may be possible to increase the ability of the bacterium to survive in large-scale culture, due to an increased ability of the cell to rapidly degrade proteins misfolded in response to the high temperatures, nonoptimal pH, and other stresses encountered during fermentor culture. Increased numbers of cells in these cultures may result in increased yields or efficiency of production of one or more desired fine chemicals, due to the relatively larger number of cells producing these compounds in the culture. Also, *C glutamicum* cells possess multiple cell-surface proteases which serve to break down external nutrients into molecules which may be more readily incorporated by the cells as carbon/energy sources or nutrients of other kinds. An increase in activity or number of these enzymes may improve this turnover and increase the levels of available nutrients, thereby improving cell growth or production. Thus, modifications of the proteases of the invention may indirectly impact *C. glutamicum* fine chemical production.

A more direct impact on fine chemical production in response to the modification of one or more of the proteases of the invention may occur when these proteases are involved in the production or degradation of a desired fine chemical. By decreasing the activity of a protease which degrades a fine chemical or a protein involved in the synthesis of a fine chemical it may be possible to increase the levels of that fine chemical (due to the decreased degradation or increased synthesis of the compound). Similarly, by increasing the activity of a protease which degrades a compound to result in a fine chemical or a protein involved in the degradation of a fine chemical, a similar result should be achieved: increased levels of the desired fine chemical from *C. glutamicum* cells containing these engineered proteins.

The aforementioned mutagenesis strategies for HA proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one skilled in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated HA nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any product produced by *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

Exemplification

Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC 13032

A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $MgSO_4 \times 7H_2O$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(NH_4)_2SO_4$, 1 g/l NaCl, 2 g/l $MgSO_4 \times 7H_2O$, 0.2 g/l $CaCl_2$, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $FeSO_4 \times H_2O$, 10 mg/l $ZnSO_4 \times 7\ H_2O$, 3 mg/l $MnCl_2 \times 4\ H_2O$, 30 mg/l H_3BO_3 20 mg/l $CoCl_2 \times 6\ H_2O$, 1 mg/l $NiCl_2 \times 6\ H_2O$, 3 mg/l $Na_2MoO_4 \times 2\ H_2O$, 500 mg/l complexing agent (EDTA or critic acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca.18 h at 37°C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20

μg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

Starting from DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (*see* e.g., Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) Proc. Natl. Acad. Sci. USA, 75:3737-3741); pACYC177 (Change & Cohen (1978) J. Bacteriol 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) Gene 53:283-286.

Example 3: DNA Sequencing and Computational Functional Analysis

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R.D. et al. (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

Example 4: *In vivo* Mutagenesis

In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain

the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) Strategies 7: 32-34.

Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J.F. *et al.* (1987) Biotechnology, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli and C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. et al. (1985) J. Bacteriol. 162:591-597, Martin J.F. et al. (1987) Biotechnology, 5:137-146 and Eikmanns, B.J. et al. (1991) Gene, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. et al. (1984) J. Bacteriol. 159306-311), electroporation (Liebl, E. et al. (1989) FEMS Microbiol. Letters, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A et al. (1990) J. Bacteriol. 172:1663-1666). It is also possible to transfer the shuttle vectors for

C. glutamicum to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) J. Mol. Biol. 166:1-19).

Example 6: Assessment of the Expression of the Mutant Protein

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E.R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

Example 7: Growth of Genetically Modified *Corynebacterium glutamicum* — Media and Culture Conditions

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al* (1989) Appl. Microbiol. Biotechnol., 32:205-210; von der Osten *et al.* (1998) Biotechnology Letters, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes, Volume II, Balows, A. *et al*. eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (*eds*. P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0

19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFC) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the microorganisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of 0.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2,5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract,

22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 8 – *In vitro* Analysis of the Function of Mutant Proteins

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) EMBO J. 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing

the medium and/or the cellular component for increased production of the desired product (i.e., an amino acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al. (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. et al. (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum*

cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biotekhnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Equivalents

Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

Table 1: Genes in the Application

Cell wall biosynthesis

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA01430	GR00417	7458	8271		N-ACETYLMURAMOYL-L-ALANINE AMIDASE (EC 3 5.1 28)
RXA02641	GR00749	5097	3022		N-ACETYLMURAMOYL-L-ALANINE AMIDASE (EC 3 5.1 28)
RXA00135	GR00021	1709	2962	BS-murA,EC-murA	UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5 1 7)
RXA02706	GR00758	6910	5813	EC-mraY,BS-mraY	PHOSPHO-N-ACETYLMURAMOYL-PENTAPEPTIDE-TRANSFERASE (EC 2 7 8 13)
RXA02702	GR00758	1572	115	EC-murC,BS-murC	UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6 3 2.8)
RXA02411	GR00703	1845	997	BS-racE,EC-murI	GLUTAMATE RACEMASE (EC 5 1 1 3)
RXA02705	GR00758	5803	4388	BS-murD	UDP-N-ACETYLMURAMOYLALANINE--D-GLUTAMATE LIGASE (EC 6 3 2 9)
RXA01254	GR00365	3807	2539		UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE (EC 6 3 2.13)
RXA02707	GR00758	7264	6920		UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D-ALANYL LIGASE (EC 6 3 2.15)
RXA02708	GR00758	7694	7260		UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D-ALANYL LIGASE (EC 6.3 2.15)
RXA02709	GR00758	8451	7723	BS-murF	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D-ALANYL LIGASE (EC 6 3 2 15)
RXA02710	GR00758	10035	8473	BS-murE,EC-murE	UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMATE--2,6-DIAMINOPIMELATE LIGASE (EC 6 3 2 13)
RXA00508	GR00127	1193	1921	BS-yncD,EC-alr	ALANINE RACEMASE (EC 5.1.1.1)
RXA01022	GR00292	3	806	EC-ddlA,BS-ddlA	D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4)
RXA02703	GR00758	2698	1610	EC-murG,BS-murG	UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE (EC 2.4 1 -)
RXA02711	GR00758	12273	10162	EC-ftsI,BS-spoVD	PENICILLIN-BINDING PROTEIN 2
RXA02859	GR10005	848	121		PENICILLIN-BINDING PROTEIN 5* PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3 4.16 4)
RXA00569	GR00152	3928	4953		PENICILLIN-BINDING PROTEIN 4
RXA00594	GR00158	3525	4457	BS-pbpF,EC-mrcA	PENICILLIN-BINDING PROTEIN 1A
RXA01828	GR00516	7736	6315	BS-pbpC	PENICILLIN-BINDING PROTEIN 3
RXA00812	GR00162	3	1187		PENICILLIN-BINDING PROTEIN 1A
RXA01510	GR00424	15370	16650		PENICILLIN-BINDING PROTEIN 4 PRECURSOR (PBP-4) (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3 4.16 4) / D-ALANYL-D-ALANINE-ENDOPEPTIDASE (EC 3 4 99 -)
RXA01608	GR00449	837	2875		(AL008883) penecillin binding protein [Mycobacterium tuberculosis]
RXA01376	GR00400	2872	3759		Glycosyltransferases, typically involved in cell wall biogenesis
RXA01270	GR00367	21652	20498	BS-yvfE,EC-b2253	perosamine synthelase

Cell division

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02704	GR00758	4382	2694	BS-spoVE,EC-ftsW	CELL DIVISIN PROTEIN FTSW
RXA02722	GR00759	2729	1404	BS-ftsZ,EC-ftsZ	CELL DIVISION PROTEIN FTSZ
RXA00009	GR00002	1545	646		CELL DIVISION PROTEIN FTSX
RXA00010	GR00002	2248	1562	BS-ftsE,EC-ftsE	CELL DIVISION ATP-BINDING PROTEIN FTSE
RXA00143	GR00022	8328	4847	EC-b2304,BS-ylhF	CELL DIVISION INHIBITOR
RXA00277	GR00043	1588	5		CELL DIVISION CONTROL PROTEIN 15 (EC 2.7 1 -)
RXA00857	GR00233	2	1291	BS-spoIIIE,EC-ftsK	CELL DIVISION PROTEIN FTSK
RXA00860	GR00234	858	1751		CELL DIVISION PROTEIN FTSK
RXA01435	GR00418	2	871		CELL DIVISION CONTROL PROTEIN 15 (EC 2 7 1 -)
RXA01511	GR00424	16655	17596		CELL CYCLE PROTEIN MESJ
RXA01513	GR00424	18388	20926	BS-ftsH,EC-hflB	CELL DIVISION PROTEIN FTSH (EC 3.4 24 -)
RXA02098	GR00630	4161	5906	BS-ftsY	CELL DIVISION PROTEIN FTSY
RXA02713	GR00758	14077	13067	BS-ylxA,EC-yabC	Hypothetical Cell Division Protein mraW
RXA02723	GR00759	3480	2984		FTSQ
RXA02651	GR00752	2041	2970	EC-sfhB,BS-ylyB	ftsH suppressor - suppresses the temperature-sensitivity of the ftsH1(Ts) mutation
RXA01426	GR00417	2777	3403	BS-gidB,EC-gidB	GLUCOSE INHIBITED DIVISION PROTEIN B
RXA02082	GR00628	14248	17154	BS-smc	CHROMOSOME SEGREGATION PROTEIN SMC2
RXA01428	GR00417	4495	5631	BS-spo0J	STAGE 0 SPORULATION PROTEIN J
RXA01640	GR00456	4661	1344	BS-yukA	STAGE III SPORULATION PROTEIN E
RXA01829	GR00518	9058	7736		STAGE V SPORULATION PROTEIN E
RXA01427	GR00417	3512	4432	BS-soj	SOJ PROTEIN
RXA01603	GR00447	14043	14683		SOJ PROTEIN

Proteolysis

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00675	GR00178	2	484		METHIONINE AMINOPEPTIDASE (EC 3 4 11 18)
RXA01609	GR00449	2740	3612	EC-map,BS-map	METHIONINE AMINOPEPTIDASE (EC 3 4 11 18)
RXA01358	GR00393	5337	6857	EC-ylfB	ATP-DEPENDENT PROTEASE LA (EC 3.4 21 53)
RXA01458	GR00420	3225	2176	BS-ylbL	ATP-DEPENDENT PROTEASE LA (EC 3 4.21.53)
RXA01654	GR00459	986	1981		(AL022121) putative alkaline serine protease [Mycobacterium tuberculosis]
RXA01868	GR00534	1640	30		ZINC METALLOPROTEASE (EC 3 4 24 -)
RXA01869	GR00534	1954	1652		ZINC METALLOPROTEASE (EC 3 4 24 -)
RXA02470	GR00715	2216	3196		ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
RXA02471	GR00715	3159	4991	BS-clpC	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA
RXA02630	GR00748	2654	1332	BS-htrA,EC-htrA	(AL021999) putative serine protease [Mycobacterium tuberculosis]
RXA02834	GR00823	3	497		ATPases with chaperone activity, ATP-dependent protease subunit

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00112	GR00016	3687	2497		PROBABLE PERIPLASMIC SERINE PROTEASE DO-LIKE PRECURSOR
RXA00566	GR00152	742	137		ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT (EC 3.4.21.92)
RXA00587	GR00152	1388	798	EC-clpP,BS-clpP	ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT (EC 3.4.21.92)
RXA01668	GR00464	2205	3920	EC-clpB	CLPB PROTEIN
RXA01120	GR00310	2349	1072	EC-clpX,BS-clpX	ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPX
RXA00744	GR00202	10722	9781		Periplasmic serine proteases
RXA00844	GR00228	3620	4453		Hypothetical Secretory Serine Protease (EC 3.4.21.-)
RXA01151	GR00324	882	5		ATP-dependent Zn proteases
RXA02317	GR00665	9664	9053		PEPTIDASE E (EC 3.4.-.-)
RXA02644	GR00751	767	117	BS-yvyE,EC-yigZ	XAA-PRO DIPEPTIDASE (EC 3.4.13.9)
RXA02820	GR00801	1	507		GAMMA-GLUTAMYLTRANSPEPTIDASE (EC 2.3.2.2)
RXA02859	GR10005	846	121		PENICILLIN-BINDING PROTEIN 5' PRECURSOR (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (EC 3.4.16.4)
RXA00137	GR00022	738	1826	BS-yqhT,EC-b2385	XAA-PRO AMINOPEPTIDASE (EC 3.4.11.9)
RXA00499	GR00125	3	959		PROLINE IMINOPEPTIDASE
RXA00877	GR00242	3	1067	EC-dcp	PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5)
RXA01014	GR00289	3	1580		AMINOPEPTIDASE N (EC 3.4.11.2)
RXA01018	GR00290	2289	3152		AMINOPEPTIDASE N (EC 3.4.11.2)
RXA01147	GR00323	1353	94		VACUOLAR AMINOPEPTIDASE I PRECURSOR (EC 3.4.11.1)
RXA01161	GR00329	1253	117		XAA-PRO AMINOPEPTIDASE (EC 3.4.11.9)
RXA01181	GR00337	1	957	BS-yuiE,EC-pepA	AMINOPEPTIDASE
RXA01277	GR00388	1738	50		PROLYL ENDOPEPTIDASE (EC 3.4.21.26)
RXA01914	GR00548	125	550		AMINOPEPTIDASE
RXA02000	GR00589	3430	3933		GAMMA-GLUTAMYLTRANSPEPTIDASE (EC 2.3.2.2)
RXA02048	GR00624	207	1580		AMINOPEPTIDASE N (EC 3.4.11.2)
RXA00821	GR00163	4075	4857		PTRB periplasmic protease
RXA00622	GR00163	4778	8193	EC-ptrB	PTRB periplasmic protease
RXA00977	GR00275	1647	2660		(L42758) proteinase [Streptomyces lividans]
RXA00982	GR00276	5194	4949		(L42758) proteinase [Streptomyces lividans]
RXA00152	GR00023	7175	5880	EC-b0489	HFLC PROTEIN (EC 3.4.-.-)
RXA02558	GR00731	4939	3965		HFLC PROTEIN (EC 3.4.-.-)
RXA00500	GR00125	989	1643		O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57)
RXA00501	GR00125	1643	2149		O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57)
RXA00502	GR00125	2156	3187	BS-ydiE,EC-ygjD	O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57)

Enzymes in general

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02384	GR00692	1355	648		ALPHA-RIBAZOLE-5'-PHOSPHATE PHOSPHATASE (EC 3.1 3 -)
RXA01115	GR00308	799	1551		3-OXOADIPATE ENOL-LACTONASE (EC 3 1 1 24)
RXA01341	GR00389	11155	9812	BS-ycnG,EC-gabT	4-AMINOBUTYRATE AMINOTRANSFERASE (EC 2.6 1 19)
RXA01728	GR00489	2452	1478	BS-patB,EC-malY	BETA C-S LYASE (EC 3 - - -) PUTATIVE AMINOTRANSFERASE
RXA02148	GR00639	16581	17703		GLYCOSYL TRANSFERASE
RXA00762	GR00204	1324	449		Acetyltransferases
RXA00897	GR00245	1	537		Acetyltransferases
RXA02214	GR00650	954	1562		Acetyltransferases
RXA02716	GR00758	16827	17387		Acetyltransferases
RXA01489	GR00424	7034	3213	BS-dhbF,EC-entF	Acetyltransferases (the isoleucine patch superfamily)
RXA01257	GR00366	1102	257	EC-yraL,BS-yabC	Predicted methyltransferases
RXA02589	GR00741	13804	13040	EC-yggH	Predicted S-adenosylmethionine-dependent methyltransferase
RXA00226	GR00032	26838	26012		SAM-dependent methyltransferases
RXA01885	GR00539	1589	2389		SAM-dependent methyltransferases
RXA02592	GR00741	18477	17707		SAM-dependent methyltransferases
RXA01795	GR00507	708	1140		MODIFICATION METHYLASE (EC 2 1.1 73)
RXA01214	GR00351	1640	3130	EC-yacK	LACCASE I PRECURSOR (EC 1.10 3 2)
RXA01250	GR00364	592	5		LACCASE I PRECURSOR (EC 1 10.3 2)
RXA02477	GR00715	10581	11201		CARBONIC ANHYDRASE (EC 4 2.1 1)
RXA00833	GR00225	374	6	EC-tpx,BS-ytgI	THIOL PEROXIDASE (EC 1 11 1 -)
RXA01224	GR00354	4186	5208	BS-yrpB	2-NITROPROPANE DIOXYGENASE (EC 1 13 11 32)
RXA01182	GR00337	1363	971		Hypothetical Oxidoreductase
RXA02531	GR00726	1228	1936		Hypothetical Oxidoreductase
RXA00689	GR00180	1401	775		BETAINE-ALDEHYDE DEHYDROGENASE PRECURSOR (EC 1 2 1.8)
RXA02192	GR00643	2	523		MORPHINE 6-DEHYDROGENASE (EC 1 1 1 218)
RXA02351	GR00679	132	1070	BS-ytnJ	NITRILOTRIACETATE MONOOXYGENASE COMPONENT A (EC 1 14.13 -)
RXA00231	GR00034	2544	1756		SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (EC 1 2 1 16)
RXA00659	GR00170	608	1228		SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (EC 1 2.1 16)
RXA01340	GR00389	9439	8450	EC-gabD,BS-ycnH	SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (EC 1 2.1 16)
RXA01498	GR00424	1598	3180		SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (EC 1 2 1 16)
RXA00787	GR00209	598	5	BS-ywmD	D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1 4 99 1)
RXA00791	GR00210	831	4		D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1 4 99 1)
RXA01057	GR00296	7548	6046		D-AMINO ACID DEHYDROGENASE LARGE SUBUNIT (EC 1 4.99.1)
RXA01055	GR00296	4821	4720		D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT (EC 1 4 99 1)
RXA01056	GR00296	5952	5053		D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT (EC 1 4 99 1)
RXA00944	GR00259	1573	602		NADPH DEHYDROGENASE 3 (EC 1.6.99.1)
RXA00389	GR00086	3	1349		oxoglutarate semialdehyde dehydrogenase (EC 1 2 1 -)

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00905	GR00247	2	694		N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3 5 1 14)
RXA00906	GR00247	630	1133		N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3 5 1 14)
RXA00907	GR00247	1143	1265		N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1.14)
RXA02101	GR00631	3104	1842		N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5 1 14)
RXA02565	GR00733	1	342		N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3 5 1 14)
RXA02567	GR00734	3	740		N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5.1 14)
RXA02855	GR10002	1693	2877	BS-hipO	N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 3.5 1.14), hippurate hydrolase
RXA00026	GR00003	3657	5042	EC-b1337	Hypothetical Amidohydrolase (EC 3.5.1.-)
RXA01971	GR00569	963	133	BS-yabD,EC-ycfH	Hypothetical Metal-Dependent Hydrolase
RXA00354	GR00068	2308	1673		Predicted hydrolases (dehalogenase-related)
RXA01802	GR00509	3461	4291	EC-b0844	Predicted hydrolases (HAD superfamily)
RXA00866	GR00236	3555	4499	BS-ykqC	Predicted Zn-dependent hydrolases
RXA02410	GR00703	792	127	BS-yhfI	Predicted Zn-dependent hydrolases
RXA00961	GR00267	2	433		SALICYLATE HYDROXYLASE (EC 1 14 13 1)
RXA00111	GR00016	930	1922	BS-yfhM	SOLUBLE EPOXIDE HYDROLASE (SEH) (EC 3 3 2 3)
RXA01932	GR00555	6479	5583		ACETYL-HYDROLASE (EC 3.1-1 -)
RXA02574	GR00739	833	1840	EC-b1107,BS-ybbD	PUTATIVE SECRETED HYDROLASE
RXA00983	GR00278	1200	4		SIALIDASE PRECURSOR (EC 3 2.1.18)
RXA00984	GR00278	1716	1300		SIALIDASE PRECURSOR (EC 3 2 1 18)
RXA02513	GR00722	93	824		SIALIDASE PRECURSOR (EC 3 2 1 18)
RXA00636	GR00167	923	1951		METAL-ACTIVATED PYRIDOXAL ENZYME
RXA00903	GR00246	637	5		Putative epimerase
RXA01090	GR00304	4035	4346		DEHYDROGENASE
RXA01224	GR00354	4186	5208	BS-yrpB	2-NITROPROPANE DIOXYGENASE (EC 1 13.11.32)
RXA01571	GR00438	1360	1959	EC-b0325	ALCOHOL DEHYDROGENASE (EC 1.1.1 1)

Genes encoding enzymes for the metabolism of inorganic compounds

N-metabolism

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01302	GR00376	370	5		NITRATE REDUCTASE ALPHA CHAIN (EC 1 7 99 4)
RXA01307	GR00377	686	6		NITRATE REDUCTASE ALPHA CHAIN (EC 1 7.99 4)
RXA01308	GR00378	1211	6	BS-narG,EC-narG	NITRATE REDUCTASE ALPHA CHAIN (EC 1 7 99 4)
RXA01309	GR00379	719	51		NITRATE REDUCTASE ALPHA CHAIN (EC 1 7 99.4)

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02017	GR00810	1731	1048		NITRATE REDUCTASE ALPHA CHAIN (EC 1.7.99.4)
RXA02018	GR00810	2788	1739	BS-narH,EC-narY	NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4)
RXA02016	GR00810	1036	260	BS-narI,EC-narI	NITRATE REDUCTASE GAMMA CHAIN (EC 1.7.99.4)
RXA00471	GR00119	2997	3686		NITRATE/NITRITE RESPONSE REGULATOR PROTEIN NARL
RXA00133	GR00021	201	1013		NITRATE/NITRITE RESPONSE REGULATOR PROTEIN NARP
RXA00650	GR00169	4017	3382	BS-yvqC,EC-narP	NITRATE/NITRITE RESPONSE REGULATOR PROTEIN NARP
RXA01189	GR00339	2545	1937	BS-yocG,EC-uhpA	NITRATE/NITRITE RESPONSE REGULATOR PROTEIN NARP
RXA01607	GR00449	123	752		NITRATE/NITRITE RESPONSE REGULATOR PROTEIN NARP
RXA00470	GR00119	1752	2951		NITRATE/NITRITE SENSOR PROTEIN NARX (EC 2.7.3.-)
RXA01589	GR00444	1033	197	BS-narA,EC-moaA	gene required for nitrate assimilation and anaerobic growth
RXA00756	GR00203	2932	1937	BS-nusA,EC-nusA	N UTILIZATION SUBSTANCE PROTEIN A
RXA00139	GR00022	2514	3224		N UTILIZATION SUBSTANCE PROTEIN B
RXA01073	GR00300	1274	2104	BS-nadE,EC-nadE	NH(3)-DEPENDENT NAD(+) SYNTHETASE (EC 6.3.5.1)
RXA01303	GR00376	1724	390	EC-narU	NITRITE EXTRUSION PROTEIN
RXA01412	GR00412	620	417		NITROGEN FIXATION PROTEIN FIXI (PROBABLE E1-E2 TYPE CATION ATPASE) (EC 3.6.1.-)
RXA00773	GR00205	3208	4350	BS-yacF,EC-yhdG	NITROGEN REGULATION PROTEIN NIFR3
RXA02748	GR00764	1	267	EC-glnK	NITROGEN REGULATORY PROTEIN P-II
RXA02745	GR00763	15350	14472	BS-yvfR	NODULATION ATP-BINDING PROTEIN I
RXA00820	GR00221	1007	1369		NODULATION PROTEIN N
RXA01059	GR00298	8782	9390	EC-b1008	OXYGEN-INSENSITIVE NAD(P)H NITROREDUCTASE (EC 1.-.-.-)
RXA01324	GR00385	3442	3741	BS-yvgZ	PLASMID PTOM9 FROM ALCALIGENES XYLOSOXIDANS NREA AND NREB GENES, COMPLETE CDS

Urease

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02264	GR00655	123	4		UREASE ALPHA SUBUNIT (EC 3.5.1.5)
RXA02274	GR00656	3	1604	BS-ureC	UREASE ALPHA SUBUNIT (EC 3.5.1.5)
RXA02265	GR00655	452	153	BS-ureA	UREASE GAMMA SUBUNIT (EC 3.5.1.5)
RXA02278	GR00656	3420	4268		UREASE OPERON URED PROTEIN
RXA02275	GR00656	1632	2102		UREASE ACCESSORY PROTEIN UREE
RXA02276	GR00656	2105	2782		UREASE ACCESSORY PROTEIN UREF
RXA02277	GR00656	2802	3416		UREASE ACCESSORY PROTEIN UREG
RXA02215	GR00650	2734	1868	EC-ygfD	Urease/hydrogenase-associated predicted GTPases

14.07.99

Phosphate and Phosphonate metabolism

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA02118	GR00638	2124	1783	EC-phnA	PHNA PROTEIN
RXA00078	GR00012	6375	5962		PHNB PROTEIN
RXA02105	GR00632	294	4		PHNB PROTEIN
RXA00912	GR00248	1772	2044		PHA(A,B,C,D,E,F,O) GENES
RXA00663	GR00173	1222	227	BS-ylaK	PHOH PROTEIN HOMOLOG
RXA00888	GR00242	14325	15341	BS-phoH,EC-b0660	PHOH PROTEIN HOMOLOG
RXA01437	GR00418	3932	2550	EC-pta,BS-pta	PHOSPHATE ACETYLTRANSFERASE (EC 2.3.1.8)
RXA00778	GR00205	9079	8246	EC-pstS	PHOSPHATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA01732	GR00491	851	1903		PHOSPHONATES-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA02877	GR10016	1467	1083		PHOSPHOPANTETHIENE PROTEIN TRANSFERASE, PPT1P
RXA01716	GR00486	384	4		EXOPOLYPHOSPHATASE (EC 3.6.1.11)
RXA02497	GR00720	10059	10985		EXOPOLYPHOSPHATASE (EC 3.6.1.11)
RXA00191	GR00029	5193	6611		ALKALINE PHOSPHATASE (EC 3.1.3.1)
RXA01477	GR00422	8489	10016	BS-phoD	ALKALINE PHOSPHATASE D PRECURSOR (EC 3.1.3.1)
RXA01509	GR00424	15169	14896	EC-ppa	INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
RXA01593	GR00447	1426	2292		4-NITROPHENYLPHOSPHATASE (EC 3.1.3.41)
RXA00100	GR00014	9512	10111	BS-ykoX	DEDA PROTEIN, similar to alkaline phosphatase
RXA00615	GR00162	3355	2774		DEDA PROTEIN
RXA02010	GR00602	79	525	EC-dedA	DEDA PROTEIN
RXA02120	GR00636	5021	4260		CARBOXYVINYL-CARBOXYPHOSPHONATE PHOSPHORYLMUTASE (EC 2.7.8.23)

Sulfate metabolism

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00072	GR00012	446	6		PHOSPHOADENOSINE PHOSPHOSULFATE REDUCTASE (EC 1.8.99.4)
RXA02548	GR00727	3	293		SULFATE ADENYLATE TRANSFERASE SUBUNIT 2 (EC 2.7.7.4)
RXA00793	GR00211	1469	2644	BS-ygcA,EC-b0935	SULFATE STARVATION-INDUCED PROTEIN 6
RXA01192	GR00342	161	733		SULFATE STARVATION-INDUCED PROTEIN 6

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA01232	GR00356	1811	2497		SULFITE OXIDASE (EC 1.8.3.1)
RXA00715	GR00188	2120	2914	EC-sseA	THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1)
RXA01664	GR00463	1306	485		THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1)
RXA02334	GR00872	2	355		THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1)
RXA00249	GR00037	8837	7884	EC-cysN	ADENYLYLSULFATE KINASE (EC 2.7.1.25)

Fe-Metabolism

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA01967	GR00567	1848	706		FERRIC ENTEROCHELIN ESTERASE HOMOLOG
RXA00070	GR00011	3436	3867	EC-fur,BS-yqfV	FERRIC UPTAKE REGULATION PROTEIN
RXA01934	GR00555	7192	7749	EC-b3279,BS-yloA	FERRIPYOCHELIN BINDING PROTEIN
RXA01997	GR00586	546	935		FERRITIN
RXA01810	GR00511	860	258		HEMIN-BINDING PERIPLASMIC PROTEIN HMUT PRECURSOR
RXA00467	GR00118	1	474		IRON REPRESSOR
RXA01082	GR00302	1486	827		IRON REPRESSOR
RXA00085	GR00013	3287	2370	EC-fecB	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA01236	GR00358	2185	1241	BS-yvrC	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA01354	GR00393	2892	1757		IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA01620	GR00451	2585	3532		IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA02052	GR00624	4588	3795	BS-yhY	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
RXA02875	GR10016	3	380	EC-bcp	bacterioferritin comigratory protein
RXA00372	GR00078	1653	2729		PERIPLASMIC-IRON-BINDING PROTEIN SHIB
RXA00088	GR00013	4389	5402		FERRIC ANGUIBACTIN-BINDING PROTEIN PRECURSOR

Mg Metabolism

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA01691	GR00474	570	4		MAGNESIUM-CHELATASE SUBUNIT CHLI
RXA01848	GR00524	1532	789		MAGNESIUM-CHELATASE SUBUNIT CHLI
RXA01849	GR00524	2004	1555		MAGNESIUM-CHELATASE SUBUNIT CHLI

Modification and degradation of aromatic compounds

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00024	GR00003	938	1882	EC-b0419,BS-yccK	ARYL-ALCOHOL DEHYDROGENASE (NADP+) (EC 1 1.1.91)
RXA02526	GR00725	4109	5314		3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE (EC 5 5.1 2)
RXA02813	GR00794	651	10		3-CARBOXY-CIS,CIS-MUCONATE CYCLOISOMERASE HOMOLOG (EC 5 5.1 2)
RXA01113	GR00307	1098	862		4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4.1.1.44)
RXA02126	GR00637	1556	1876		4-CARBOXYMUCONOLACTONE DECARBOXYLASE (EC 4 1 1.44)
RXA01465	GR00421	4121	2961	BS-ykfB	MUCONATE CYCLOISOMERASE (EC 5 5 1 1)
RXA02316	GR00665	9038	8025		MUCONATE CYCLOISOMERASE (EC 5 5 1 1)
RXA01464	GR00421	2945	2655		MUCONOLACTONE ISOMERASE (EC 5.3 3.4)
RXA02603	GR00742	7742	8737		4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2 5 1 -)
RXA02839	GR00832	3	419	EC-ubiA	4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)
RXA02674	GR00753	15614	14163	BS-yfmT	BENZALDEHYDE DEHYDROGENASE (NAD+) (EC 1 2.1.28)
RXA01502	GR00424	8385	9617	EC-b2542,BS-nasD	BENZENE 1,2-DIOXYGENASE SYSTEM FERREDOXIN--NAD(+) REDUCTASE COMPONENT (EC 1 18.1.3)
RXA02828	GR00813	15	572	BS-yfiE	BIPHENYL-2,3-DIOL 1,2-DIOXYGENASE III (EC 1.13.11.39)
RXA01918	GR00549	4644	5057		CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2.1 1 104)
RXA02064	GR00626	5223	4585		CAFFEOYL-COA O-METHYLTRANSFERASE (EC 2 1 1 104)
RXA00639	GR00168	665	6		CATECHOL 1,2-DIOXYGENASE (EC 1 13 11.1)
RXA00797	GR00212	445	804		DIBENZOTHIOPHENE DESULFURIZATION ENZYME A
RXA01653	GR00458	1909	971		DIBENZOTHIOPHENE DESULFURIZATION ENZYME A
RXA00792	GR00211	159	1355		DIBENZOTHIOPHENE DESULFURIZATION ENZYME C
RXA02530	GR00726	20	469		DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 1 (EC 1 14 13 8)
RXA02083	GR00629	1720	311		DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 2 (EC 1 14 13.8)
RXA00892	GR00243	2188	1295		PARANITROBENZYL ESTERASE (EC 3.1.1.-)
RXA02092	GR00629	12153	10516	BS-pnbA	PARANITROBENZYL ESTERASE (EC 3 1 1 -)
RXA00658	GR00170	321	4		PHENOL 2 MONOOXYGENASE (EC 1 14 13 7)
RXA01385	GR00406	5320	3440		PHENOL 2 MONOOXYGENASE (EC 1.14.13.7)
RXA01461	GR00421	463	5		PROTOCATECHUATE 3,4-DIOXYGENASE ALPHA CHAIN (EC 1 13.11.3)
RXA01462	GR00421	1167	478		PROTOCATECHUATE 3,4-DIOXYGENASE BETA CHAIN (EC 1.13.11.3)
RXA00640	GR00168	1083	1331		TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1 14 12 -)
RXA00641	GR00168	1533	2573		TOLUATE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12 -)
RXA00642	GR00168	2616	3107		TOLUATE 1,2-DIOXYGENASE BETA SUBUNIT (EC 1 14.12.-)

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00843	GR00168	3122	4657		TOLUATE 1,2-DIOXYGENASE ELECTRON TRANSFER COMPONENT
RXA01993	GR00584	1	386		VANILLATE DEMETHYLASE (EC 1 14.-.-)
RXA02012	GR00604	2	670	EC-b1803	VANILLATE DEMETHYLASE (EC 1.14.-.-)
RXA01994	GR00584	373	1347		VANILLATE DEMETHYLASE OXIDOREDUCTASE (EC 1.-.-.-)
RXA02535	GR00726	6599	7753		XYLENE MONOOXYGENASE ELECTRON TRANSFER COMPONENT
RXA00964	GR00269	1575	451		1-hydroxy-2-naphthoate 1,2-dioxygenase (EC 1.13 11.38)
RXA01466	GR00422	826	5		ARYLESTERASE (EC 3 1 1 2)
RXA02895	GR10037	671	6		CHLOROCATECHOL 1,2-DIOXYGENASE (EC 1 13 11 1)
RXA02449	GR00710	1458	2360		hydroxyquinol 1,2-dioxygenase (EC 1 13 11 37)
RXA00178	GR00028	304	1188		HYDROXYQUINOL-1, 2-DIOXYGENASE
RXA02111	GR00632	4310	5593	BS-nadA	QUINOLINATE SYNTHETASE A
RXA00644	GR00168	4657	5484		CIS-1,2-DIHYDROXYCYCLOHEXA-3,5-DIENE-1-CARBOXYLATE DEHYDROGENASE (EC 1 3 1 55)
RXA00177	GR00028	3	290		MALEYLACETATE REDUCTASE (EC 1 3 1 32) metabolism of 2,4,5-trichlorophenoxyacetic acid
RXA02448	GR00710	340	1428	EC-fucO	MALEYLACETATE REDUCTASE (EC 1.3.1 32)
RXA00048	GR00008	2185	527	EC-b0347,BS-yhjG	3-(3-HYDROXYPHENYL) PROPIONATE HYDROXYLASE
RXA01126	GR00313	2	565	BS-yisK,EC-b1180	POSSIBLE 2-HYDROXYHEPTA-2,4-DIENE-1,7- DIOATE ISOMAERASE
RXA01116	GR00309	949	338	BS-yxjE,EC-atoA	SUCCINYL-COA 3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (EC 2.8.3 5)
RXA01117	GR00309	1713	973	BS-yxjD,EC-atoD	SUCCINYL-COA 3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (EC 2 8 3 5)
RXA00772	GR00205	2715	1210	EC-b2920	SUCCINYL-COA COENZYME A TRANSFERASE (EC 2 8 3 -)
RXA01288	GR00372	2018	1644		SUCCINYL-COA:COENZYME A TRANSFERASE (EC 2 8.3.-)
RXA01115	GR00308	799	1551		3-OXOADIPATE ENOL-LACTONASE (EC 3.1.1 24)
RXA00685	GR00179	7751	7269		CYTOCHROME P450 116 (EC 1.14.-.-)

Modification and degradation of aliphatic compounds

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00289	GR00048	7376	6633	BS-yvbT,EC-yhbW	ALKANAL MONOOXYGENASE ALPHA CHAIN (EC 1 14 14 3)
RXA00332	GR00057	16086	15385		ALKANAL MONOOXYGENASE ALPHA CHAIN (EC 1 14 14.3)
RXA01838	GR00519	2	820		ALKANAL MONOOXYGENASE ALPHA CHAIN (EC 1.14 14 3)
RXA02643	GR00750	1603	560		ALKANAL MONOOXYGENASE ALPHA CHAIN (EC 1 14 14 3)
RXA01933	GR00555	6590	7192		2-HALOALKANOIC ACID DEHALOGENASE I (EC 3 8 1 2)
RXA02351	GR00679	132	1070	BS-ylnJ	NITRILOTRIACETATE MONOOXYGENASE COMPONENT A (EC 1.14 13.-)

14.07.99

TABLE 2: GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvat corboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 03/21/90
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moeckel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent: WO 9519442-A 5 07/20/95
AB003132	murC, ftsQ; ftsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from coryneform bacteria," *Biochem. Biophys. Res. Commun.*, 236(2):383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from Coryneform bacteria," *Appl. Microbiol. Biotechnol.*, 51(2):223-228 (1999)
AB018530	dtsR		Kimura, E. et al. "Molecular cloning of a novel gene, dtsR, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*," *Biosci. Biotechnol. Biochem.*, 60(10):1565-1570 (1996)
AB018531	dtsR1; dtsR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	gltB; gltD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamolytransferase	
AF036932	aroD	3-dehydroquinate dehydratase	

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF038548	pyc	Pyruvate carboxylase	
AF038651	dciAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	Wehmeier, L. et al. "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp metabolism," *Microbiology*, 144:1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase, ornithine acetyltransferase; N-acetylglutamate kinase, acetylornithine transminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in Corynebacterium glutamicum," *Mol Cells.*, 8(3):286-294 (1998)
AF053071	aroB	Dehydroquinate synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Dusch, N. et al. "Expression of the Corynebacterium glutamicum panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in Escherichia coli," *Appl. Environ. Microbiol.*, 65(4)1530-1539 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquinate synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete)	Wehrmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with Corynebacterium glutamicum," *J. Bacteriol.*, 180(12):3159-3165 (1998)
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase; ?; high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD; srp; amtP	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzmye); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in Corynebacterium glutamicum; Isolation of genes involved in biochemical characterization of corresponding proteins," *FEMS Microbiol.*, 173(2):303-310 (1999)
AJ132968	cat	Chloramphenicol acetyl transferase	
AJ224946	mqo	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from Corynebacterium glutamicum," *Eur. J. Biochem.*, 254(2):395-403 (1998)
AJ238250	ndh	NADH dehydrogenase	
AJ238703	porA	Porin	Lichtinger, T. et al. "Biochemical and biophysical characterization of the cell wall porin of Corynebacterium glutamicum: The channel is formed by a low molecular mass polypeptide," *Biochemistry*, 37(43):15024-15032 (1998)
D17429		Transposable element IS31831	Vertes, A.A. et al. "Isolation and characterization of IS31831, a transposable element from Corynebacterium glutamicum," *Mol. Microbiol.*, 11(4):739-746 (1994)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the Corynebacterium glutamicum (Brevibacterium lactofermentum AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," *Microbiology*, 142:3347-3354 (1996)
E01358	hdh, hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-thereonine and L-isoleucine," Patent: JP 1987232392-A 1 10/12/87
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-thereonine and L-isoleucine," Patent: JP 1987232392-A 2 10/12/87
E01375		Tryptophan operon	
E01376	trpL; trpE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E03937		Biotin-synthase	Hatakeyama, K. et al "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 10/02/92
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04307		Flavum aspartase	Kurusu, Y. et al "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 02/09/93
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04484		Prephenate dehydratase	Sotouchi, N. et al "Production of L-phenylalanine by fermentation," Patent: JP 1993076352-A 2 03/30/93
E05108		Aspartokinase	Fugono, N et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 07/27/93
E05112		Dihydro-dipichorinate synthetase	Hatakeyama, K. et al "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 07/27/93

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent. JP 1993284970-A 1 11/02/93
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent. JP 1993284972-A 1 11/02/93
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent. JP 1993344881-A 1 12/27/93
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent JP 1993344893-A 1 12/27/93
E06825		Aspartokinase	Sugimoto, M. et al "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membraneous protein to membrane," Patent: JP 1994169780-A 1 06/21/94
E08177		Aspartokinase	Sato, Y et al "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08178, E08179, E08180, E08181, E08182		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08232		Acetohydroxy-acid isomeroreductase	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroreductase," Patent. JP 1994277067-A 1 10/04/94
E08234	secE		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 10/04/94
E08643		FT aminotransferase and desthiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08649		Aspartase	Kohama, K. et al "DNA fragment having promoter function in coryneform bacterium," Patent. JP 1995031478-A 1 02/03/95
E08900		Dihydrodipicolinate reductase	Madori, M. et al "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 03/20/95
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent. JP 1995075579-A 1 03/20/95
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-trypophan," Patent: JP 1997028391-A 1 02/04/97
E12760, E12759, E12758		transposase	Moriya, M. et al "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12764		Arginyl-tRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12767		Dihydrodipicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12773		Dihydrodipicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent: JP 1997224661-A 1 09/02/97
L01508	IlvA	Threonine dehydratase	Moeckel, B. et al. "Functional and structural analysis of the threonine dehydratase of Corynebacterium glutamicum," *J. Bacteriol.*, 174:8065-8072 (1992)
L07603	EC 4.2.1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C et al. "The cloning and nucleotide sequence of Corynebacterium glutamicum 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," *FEMS Microbiol Lett.*, 107:223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit; Acetohydroxy acid isomeroreductase	Keilhauer, C. et al. "Isoleucine synthesis in Corynebacterium glutamicum: molecular analysis of the ilvB-ilvN-ilvC operon," *J. Bacteriol.*, 175(17):5595-5603 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
L18874	PtsM	Phosphoenolpyruvate sugar phosphotransferase	Fouet, A et al "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system expression in Escherichia coli and homology to enzymes II from enteric bacteria," *PNAS USA*, 84(24) 8773-8777 (1987); Lee, J.K. et al. "Nucleotide sequence of the gene encoding the Corynebacterium glutamicum mannose enzyme II and analyses of the deduced protein sequence," *FEMS Microbiol Lett*, 119(1-2).137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium glutamicum," *J Microbiol. Biotechnol*, 4(4) 256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum," *Appl. Environ Microbiol.*, 60(7):2501-2507 (1994)
L28760	aceA	Isocitrate lyase	
L35906	dtxr	Diphtheria toxin repressor	Oguiza, J.A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the Corynebacterium diphtheriae dtxR from Brevibacterium lactofermentum," *J. Bacteriol*, 177(2):465-467 (1995)
M13774		Prephenate dehydratase	Follettie, M.T. et al. "Molecular cloning and nucleotide sequence of the Corynebacterium glutamicum pheA gene," *J Bacteriol.*, 167:695-702 (1986)
M16175	5S rRNA		Park, Y-H et al. "Phylogenetic analysis of the coryneform bacteria by 5S rRNA sequences," *J Bacteriol*, 169:1801-1806 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52.191-200 (1987)
M16664	trpA	Tryptophan synthase, 3'end	Sano, K. et al "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52:191-200 (1987)
M25819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of Corynebacterium glutamicum ATCC13032," *Gene*, 77(2):237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen Microbiol*, 138.1167-1175 (1992)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
M85107, M85108		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen Microbiol*, 138:1167-1175 (1992)
M89931	aecD; brnQ, yhbw	Beta C-S lyase, branched-chain amino acid uptake carrier, hypothetical protein yhbw	Rossol, I. et al. "The Corynebacterium glutamicum aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," *J. Bacteriol.*, 174(9):2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the brnQ gene product," *Arch. Microbiol.*, 169(4):303-312 (1998)
S59299	trp	Leader gene (promoter)	Herry, D.M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: identification of a mutation in the trp leader sequence," *Appl. Environ. Microbiol.*, 59(3):791-799 (1993)
U11545	trpD	Anthranilate phosphoribosyltransferase	O'Gara, J.P. and Dunican, L.K. (1994) Complete nucleotide sequence of the Corynebacterium glutamicum ATCC 21850 trpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR, clgIIR	Putative type II 5-cytosoine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al. "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from Corynebacterium glutamicum ATCC 13032 and analysis of its role in intergeneric conjugation with Escherichia coli," *J. Bacteriol.*, 176(23):7309-7319 (1994); Schafer, A. et al. "The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine in an McrBC-deficient Escherichia coli strain," *Gene*, 203(2):95-101 (1997)
U14965	recA		
U31224	ppx		Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)
U31230	obg; proB, unkdh	?;gamma glutamyl kinase;similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J. Bacteriol.*, 178(15):4412-4419 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
U31281	bioB	Biotin synthase	Serebriiskii, I.G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of Methylobacillus flagellatum and Corynebacterium glutamicum," *Gene,* 175:15-22 (1996)
U35023	thtR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A Corynebacterium glutamicum gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," *Arch Microbiol*, 166(2):76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A Corynebacterium glutamicum gene conferring multidrug resistance in the heterologous host Escherichia coli," *J Bacteriol*, 179(7):2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'5''-aminoglycoside phosphotransferase	
U89648		Corynebacterium glutamicum unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the Brevibacterium lactofermentum tryptophan operon," *Nucleic Acids Res*, 14(24):10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al "Nucleic sequence of the lysA gene of Corynebacterium glutamicum and possible mechanisms for modulation of its expression," *Mol Gen. Genet*, 212(1):112-119 (1988)
X14234	EC 4.1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B.J. et al. "The Phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: Molecular cloning, nucleotide sequence, and expression," *Mol Gen. Genet*, 218(2):330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," *Plant Mol Biol*, 21 (3):487-502 (1993)
X17313	fda	Fructose-bisphosphate aldolase	Von der Osten, C.H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the Corynebacterium glutamicum fda gene: structural comparison of C. glutamicum fructose-1, 6-biphosphate aldolase to class I and class II aldolases," *Mol Microbiol*,
X53993	dapA	L-2, 3-dihydrodipicolinate synthetase (EC 4.2.1.52)	Bonnassie, S. et al "Nucleic sequence of the dapA gene from Corynebacterium glutamicum," *Nucleic Acids Res*, 18(21):6421 (1990)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum, and the attP site of lambdacorynephage," *FEMS Microbiol. Lett.*, 66:299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the Corynebacterium glutamicum lysA gene," *Mol. Microbiol.*, 4(11):1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component 1	Heery, D.M. et al. "Nucleotide sequence of the Corynebacterium glutamicum trpE gene," *Nucleic Acids Res.*, 18(23):7138 (1990)
X56037	thrC	Threonine synthase	Han, K.S. et al. "The molecular structure of the Corynebacterium glutamicum threonine synthase gene," *Mol. Microbiol.*, 4(10):1693-1702 (1990)
X56075	attB-related site	Attachment site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum, and the attP site of lambdacorynephage," *FEMS Microbiol. Lett.*, 66:299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit, Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from Corynebacterium glutamicum," *Mol. Microbiol.*, 5(5):1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspertate beta-semialdehyde dehydrogenase gene asd in Corynebacterium glutamicum," *Mol. Gen. Genet.*, 224(3):317-324 (1990)
X59403	gap,pgk; tpi	Glyceraldehyde-3-phosphate; phosphoglycerate kinase, triosephosphate isomerase	Eikmanns, B.J. "Identification, sequence analysis, and expression of a Corynebacterium glutamicum gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomeras," *J. Bacteriol.*, 174(19):6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E.R. et al. "Molecular analysis of the Corynebacterium glutamicum gdh gene encoding glutamate dehydrogenase," *Mol. Microbiol.*, 6(3):317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A.H. et al. "Molecular analysis of the Corynebacterium glutamicum lysI gene involved in lysine uptake," *Mol. Microbiol.*, 5(12):2995-3005 (1991)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X66078	cop1	Ps1 protein	Joliff, G. et al "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of Corynebacterium glutamicum: The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," *Mol. Microbiol.*, 6(16):2349-2362 (1992)
X66112	glt	Citrate synthase	Eikmanns, B.J. et al. "Cloning sequence, expression and transcriptional analysis of the Corynebacterium glutamicum gltA gene encoding citrate synthase," *Microbiol*, 140:1817-1828 (1994)
X67737	dapB	Dihydrodipicolinate reductase	
X69103	csp2	Surface layer protein PS2	Peyret, J.L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in Corynebacterium glutamicum," *Mol Microbiol*, 9(1):97-109 (1993)
X69104		IS3 related insertion element	Bonamy, C et al. "Identification of IS1206, a Corynebacterium glutamicum IS3-related insertion sequence and phylogenetic analysis," *Mol. Microbiol.*, 14(3):571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in Corynebacterium glutamicum: enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," *Appl Environ. Microbiol*, 60(1):133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B.J. et al "Cloning sequence analysis, expression, and inactivation of the Corynebacterium glutamicum icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," *J. Bacteriol*, 177(3):774-782 (1995)
X72855	GDHA	Glutamate dehydrogenase (NADP+)	
X75083, X70584	mtrA	5-methyltryptophan resistance	Heery, D.M. et al. "A sequence from a tryptophan-hyperproducing strain of Corynebacterium glutamicum encoding resistance to 5-methyltryptophan," *Biochem Biophys Res. Commun*, 201(3):1255-1262 (1994)
X75085	recA		Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of Corynebacterium glutamicum and Brevibacterium lactofermentum," *Appl Microbiol Biotechnol*, 42(4):575-580 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Reinscheid, D.J. et al. "Characterization of the isocitrate lyase gene from Corynebacterium glutamicum and biochemical analysis of the enzyme," *J Bacteriol*, 176(12):3474-3483 (1994)
X76875		ATPase beta-subunit	Ludwig, W et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from Corynebacterium glutamicum," *DNA Seq.*, 4(6):403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D.J. et al. "Malate synthase from Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase: sequence analysis," *Microbiology*, 140:3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F.A. et al. "Phylogenetic analysis of the genera Rhodococcus and Norcardia and evidence for the evolutionary origin of the genus Norcardia from within the radiation of Rhodococcus species," *Microbiol.*, 141:523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronemeyer, W. et al "Structure of the gluABCD cluster encoding the glutamate uptake system of Corynebacterium glutamicum," *J. Bacteriol.*, 177(5):1152-1158 (1995)
X81379	dapE	Succinyldiaminopimelate desuccinylase	Wehrmann, A. et al "Analysis of different DNA fragments of Corynebacterium glutamicum complementing dapE of Escherichia coli," *Microbiology*, 40:3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R. et al. "Phylogeny of the genus Corynebacterium deduced from analyses of small-subunit ribosomal DNA sequences," *Int. J. Syst. Bacteriol.*, 45(4):740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol.*, 177(24):7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol.*, 177(24):7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences," *Int. J. Syst. Bacteriol.*, 45(4):724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehrmann, A. et al. "Functional analysis of sequences adjacent to dapE of Corynebacterium glutamicumproline reveals the presence of aroP, which encodes the aromatic amino acid transporter," *J. Bacteriol.*, 177(20):5991-5993 (1995)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X86157	argB, argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway," *Microbiology*, 142:99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase, acetate kinase	Reinscheid, D.J. et al. "Cloning, sequence analysis, expression and inactivation of the Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase and acetate kinase," *Microbiology*, 145:503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," *J. Bacteriol.*, 178(7):1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90363		Promoter fragment F45	Patek, M et al "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R.M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of Corynebacterium glutamicum," *J Biol Chem*, 271(10):5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H et al. "Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine," *J Bacteriol.*, 178(17):5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of Corynebacterium glutamicum, encoding two enzymes involved in L-lysine synthesis," *Biotechnol Lett.*, 19:1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein, Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum," *Mol Microbiol.*, 22(5):815-826 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X96580	panB, panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-alanine ligase; xylulokinase	Sahm, H et al. "D-pantothenate synthesis in Corynebacterium glutamicum and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," *Appl Environ Microbiol*, 65(5).1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	
X99289		Elongation factor P	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer Brevibacterium lactofermentum (Corynebacterium glutamicum ATCC 13869)," *Gene*, 198:217-222 (1997)
Y00140	thrB	Homoserine kinase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res*, 15(9).3922 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from Corynebacterium glutamicum," *Nucleic Acids Res*, 15(9):3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L.M. et al "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(24):10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O.P. et al. "Nucleotide sequence and fine structural analysis of the Corynebacterium glutamicum hom-thrB operon," *Mol Microbiol.*, 2(1):63-72 (1988)
Y08964	murC, ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase, division initiation protein or cell division protein; cell division protein	Honrubia, M.P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from Brevibacterium lactofermentum," *Mol Gen Genet.*, 259(1):97-104 (1998)
Y09163	putP	High affinity proline transport system	Peter, H. et al. "Isolation of the putP gene of Corynebacterium glutamicumproline and characterization of a low-affinity uptake system for compatible solutes," *Arch Microbiol.*, 168(2):143-151 (1997)
Y09548	pyc	Pyruvate carboxylase	Peters-Wendisch, P.G. et al. "Pyruvate carboxylase from Corynebacterium glutamicum: characterization, expression and inactivation of the pyc gene," *Microbiology*, 144.915-927 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Patek, M. et al. "Analysis of the leuB gene from Corynebacterium glutamicum," *Appl Microbiol. Biotechnol.*, 50(1):42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of corynephage Phi-16: The construction of an integration vector," *Microbiol.*, 145:539-548 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of Corynebacterium glutamicum glnA gene encoding glutamine synthetase I," *FEMS Microbiol Lett*, 154(1):81-88 (1997)
Y16642	lpd	Dihydrolipoamide dehydrogenase	
Y18059		Attachment site Corynephage 304L	Moreau, S. et al "Analysis of the integration functions of φ304L: An integrase module among corynephages," *Virology*, 255(1):150-159 (1999)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminopimelate decarboxylase (partial)	Oguiza, J.A. et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum. Regulation of argS-lysA cluster expression by arginine," *J Bacteriol*, 175(22):7356-7362 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of Brevibacterium lactofermentum encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," *J. Bacteriol*, 175(9):2743-2749 (1993)
Z29563	thrC	Threonine synthase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," *Appl Environ Microbiol*, 60(7)2209-2219 (1994)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	
Z49822	sigA	SigA sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J Bacteriol*, 178(2):550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Oguiza, J.A. et al "The galE gene encoding the UDP-galactose 4-epimerase of Brevibacterium lactofermentum is coupled transcriptionally to the dmdR gene," *Gene*, 177:103-107 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol*, 178(2):550-553 (1996)
Z66534		Transposase	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of Brevibacterium lactofermentum ATCC 13869," *Gene*, 170(1):91-94 (1996)

[1] A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

TABLE 3: Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
Brevibacterium	ammoniagenes	21054							
Brevibacterium	ammoniagenes	19350							
Brevibacterium	ammoniagenes	19351							
Brevibacterium	ammoniagenes	19352							
Brevibacterium	ammoniagenes	19353							
Brevibacterium	ammoniagenes	19354							
Brevibacterium	ammoniagenes	19355							
Brevibacterium	ammoniagenes	19356							
Brevibacterium	ammoniagenes	21055							
Brevibacterium	ammoniagenes	21077							
Brevibacterium	ammoniagenes	21553							
Brevibacterium	ammoniagenes	21580							
Brevibacterium	ammoniagenes	39101							
Brevibacterium	butanicum	21196							
Brevibacterium	divaricatum	21792	P928						
Brevibacterium	flavum	21474							
Brevibacterium	flavum	21129							
Brevibacterium	flavum	21518							
Brevibacterium	flavum			B11474					
Brevibacterium	flavum			B11472					
Brevibacterium	flavum	21127							
Brevibacterium	flavum	21128							
Brevibacterium	flavum	21427							
Brevibacterium	flavum	21475							
Brevibacterium	flavum	21517							
Brevibacterium	flavum	21528							
Brevibacterium	flavum	21529							
Brevibacterium	flavum			B11477					

14.07.99

Brevibacterium	flavum			B11478					
Brevibacterium	flavum	21127							
Brevibacterium	flavum			B11474					
Brevibacterium	healii	15527							
Brevibacterium	ketoglutamicum	21004							
Brevibacterium	ketoglutamicum	21089							
Brevibacterium	ketosoreductum	21914							
Brevibacterium	lactofermentum				70				
Brevibacterium	lactofermentum				74				
Brevibacterium	lactofermentum				77				
Brevibacterium	lactofermentum	21798							
Brevibacterium	lactofermentum	21799							
Brevibacterium	lactofermentum	21800							
Brevibacterium	lactofermentum	21801							
Brevibacterium	lactofermentum			B11470					
Brevibacterium	lactofermentum			B11471					
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	21420							
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	31269							
Brevibacterium	linens	9174							
Brevibacterium	linens	19391							
Brevibacterium	linens	8377							
Brevibacterium	paraffinolyticum					11160			
Brevibacterium	spec.						717.73		
Brevibacterium	spec.						717.73		
Brevibacterium	spec.	14604							
Brevibacterium	spec.	21860							
Brevibacterium	spec.	21864							
Brevibacterium	spec.	21865							
Brevibacterium	spec.	21866							
Brevibacterium	spec.	19240							

Corynebacterium	acetoacidophilum	21476							
Corynebacterium	acetoacidophilum	13870							
Corynebacterium	acetoglutamicum			B11473					
Corynebacterium	acetoglutamicum			B11475					
Corynebacterium	acetoglutamicum	15806							
Corynebacterium	acetoglutamicum	21491							
Corynebacterium	acetoglutamicum	31270							
Corynebacterium	acetophilum			B3671					
Corynebacterium	ammoniagenes	6872						2399	
Corynebacterium	ammoniagenes	15511							
Corynebacterium	fujiokense	21496							
Corynebacterium	glutamicum	14067							
Corynebacterium	glutamicum	39137							
Corynebacterium	glutamicum	21254							
Corynebacterium	glutamicum	21255							
Corynebacterium	glutamicum	31830							
Corynebacterium	glutamicum	13032							
Corynebacterium	glutamicum	14305							
Corynebacterium	glutamicum	15455							
Corynebacterium	glutamicum	13058							
Corynebacterium	glutamicum	13059							
Corynebacterium	glutamicum	13060							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum	21513							
Corynebacterium	glutamicum	21526							
Corynebacterium	glutamicum	21543							
Corynebacterium	glutamicum	13287							
Corynebacterium	glutamicum	21851							
Corynebacterium	glutamicum	21253							
Corynebacterium	glutamicum	21514							
Corynebacterium	glutamicum	21516							
Corynebacterium	glutamicum	21299							

Corynebacterium	glutamicum	21300							
Corynebacterium	glutamicum	39684							
Corynebacterium	glutamicum	21488							
Corynebacterium	glutamicum	21649							
Corynebacterium	glutamicum	21650							
Corynebacterium	glutamicum	19223							
Corynebacterium	glutamicum	13869							
Corynebacterium	glutamicum	21157							
Corynebacterium	glutamicum	21158							
Corynebacterium	glutamicum	21159							
Corynebacterium	glutamicum	21355							
Corynebacterium	glutamicum	31808							
Corynebacterium	glutamicum	21674							
Corynebacterium	glutamicum	21562							
Corynebacterium	glutamicum	21563							
Corynebacterium	glutamicum	21564							
Corynebacterium	glutamicum	21565							
Corynebacterium	glutamicum	21566							
Corynebacterium	glutamicum	21567							
Corynebacterium	glutamicum	21568							
Corynebacterium	glutamicum	21569							
Corynebacterium	glutamicum	21570							
Corynebacterium	glutamicum	21571							
Corynebacterium	glutamicum	21572							
Corynebacterium	glutamicum	21573							
Corynebacterium	glutamicum	21579							
Corynebacterium	glutamicum	19049							
Corynebacterium	glutamicum	19050							
Corynebacterium	glutamicum	19051							
Corynebacterium	glutamicum	19052							
Corynebacterium	glutamicum	19053							
Corynebacterium	glutamicum	19054							

14.07.99

Corynebacterium	glutamicum	19055							
Corynebacterium	glutamicum	19056							
Corynebacterium	glutamicum	19057							
Corynebacterium	glutamicum	19058							
Corynebacterium	glutamicum	19059							
Corynebacterium	glutamicum	19060							
Corynebacterium	glutamicum	19185							
Corynebacterium	glutamicum	13286							
Corynebacterium	glutamicum	21515							
Corynebacterium	glutamicum	21527							
Corynebacterium	glutamicum	21544							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum			B8183					
Corynebacterium	glutamicum			B8182					
Corynebacterium	glutamicum			B12416					
Corynebacterium	glutamicum			B12417					
Corynebacterium	glutamicum			B12418					
Corynebacterium	glutamicum			B11476					
Corynebacterium	glutamicum	21608							
Corynebacterium	lilium		P973						
Corynebacterium	nitrilophilus	21419				11594			
Corynebacterium	spec.		P4445						
Corynebacterium	spec.		P4446						
Corynebacterium	spec.	31088							
Corynebacterium	spec.	31089							
Corynebacterium	spec.	31090							
Corynebacterium	spec.	31090							
Corynebacterium	spec	31090							20145
Corynebacterium	spec.	15954							
Corynebacterium	spec.	21857							
Corynebacterium	spec.	21862							
Corynebacterium	spec.	21863							

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Coleccion Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4th edn), World federation for culture collections world data center on microorganisms, Saimata, Japen.

>>RXA02351-amino acid sequence
(1-939, translated) 313 residues
VEGSVEKLGL ISWWEELART AERGKLDAVF LADGQAINPV GLENGPGWFL EPVTALTAMA RATNNIGLIS
TISSTFWQPF HAARMIASLD HISGGRAGIN VVTSMTDAEA RNHGMDALPG HDVRYARAAE FIETITALWD
SWPAESLVMD RAGKFADSSL IKSIDHDGEF FQVAGPLNIP SPPQGRPVLF QAGSSPQGRE IAAKYAEAIY
SVAWDLEQAQ DYRSDIHARA TAQGREPMPV LPGLVTFVGT TVEEARAKQQ ALNALLPVKD SLNQLSFFVG
QDCSTWDLDA PPPPLPPLEE FSGPKGRYET VLR
>RXA02351-nucleotide sequence A: upstream
TGACACTTTACAGACTGGTTTTCAACTAATGACACCGAAAGAAATACACCTCAACCTTTTTGCTTTCGGTGCCGGGC
ACCACGCGGCGGCGTGGCGAGCG
>RXA02351-nucleotide sequence B: coding region
GTGGAGGGAAGCGTCGAAAAGCTGGGTTTAATTTCCTGGTGGGAGGAACTCGCGCGCACCGCTGAGCGGGGCAAGCT
GGATGCGGTCTTTTTGGCCGATGGGCAGGCGATTAATCCGGTCGGTCTGGAGAATGGGCCGGGCTGGTTTTTGGAGC
CGGTGACCGCGTTGACTGCGATGGCGCGGGCGACGAACAATATTGGGTTGATCAGCACAATTTCCAGTACGTTTTGG
CAGCCGTTTCATGCGGCGCGGATGATCGCCAGCTTGGATCATATTTCGGGTGGGCGTGCTGGAATCAATGTGGTGAC
ATCGATGACCGATGCGGAGGCGCGTAACCACGGGATGGATGCGTTGCCGGGTCACGATGTTCGCTATGCGCGCGCTG
CGGAATTTATTGAAACCATCACTGCGCTGTGGGATTCTTGGCCTGCGGAAAGTTTGGTGATGGATCGTGCTGGAAAA
TTTGCGGACTCCTCGCTCATTAAATCTATCGATCATGATGGTGAGTTCTTCCAAGTCGCTGGTCCGCTGAATATCCC
CAGTCCTCCGCAGGGTCGACCCGTACTTTTTCAGGCTGGATCCTCACCGCAAGGACGGGAAATCGCTGCGAAATACG
CCGAGGCAATTTACTCTGTGGCGTGGGATTTGGAGCAAGCGCAAGATTATCGCTCTGATATTCATGCTCGTGCCACT
GCCCAGGGTCGCGAGCCCATGCCGGTGCTTCCTGGTTTGGTGACTTTTGTTGGCACGACCGTGGAAGAAGCGCGTGC
AAAACAGCAGGCTCTTAATGCGTTGCTGCCGGTCAAAGACTCACTAAATCAGTTGAGTTTCTTTGTGGGTCAAGATT
GCTCGACGTGGGATTTGGATGCACCTCCCCCACCACTGCCACCGCTAGAAGAGTTTTCCGGTCCTAAAGGCAGGTAC
GAAACGGTCCTGCGG

>>RXA02531-amino acid sequence
(1-711, translated) 237 residues
MKIPLYMAEE AAALNLLADG RLALGVSRGS PEPAEKGWEA FGYDGGDDPK AAGMAREKFL RFLDAIDGRP
MSIASENQYP RLYHPGTPLP IFPHDLDLGK SIWWGAGSHN TAEQAARDGV NLMSSTLVAE ATGQSFGDLQ
ADQIAFYRQA WKEAGHDWTP RVSVSRSIFP IVTDRDRELF GLQGQGGDQV GILDDTRSTF GRSYAGSPDE
LIDQLQGRQS CDGSRHLDAH RPQPNGC
>RXA02531-nucleotide sequence A: upstream
CACTTCGCTCCCCAAGGTACATCCCCGATGCCACTTCTTGGAGCCATCATCGGTGCCACCAAACACATTGAAGTGGG
CACTGGAGTAGTGGATATGCGTT
>RXA02531-nucleotide sequence B: coding region
ATGAAAATCCCTTTGTATATGGCCGAGGAAGCAGCTGCTCTCAATCTGCTTGCCGACGGCCGACTAGCCCTCGGAGT
TTCCAGGGGATCACCCGAACCAGCCGAGAAGGGTTGGGAAGCTTTCGGCTACGACGGCGGTGATGATCCTAAAGCTG
CAGGCATGGCACGGGAGAAATTCCTTCGCTTCCTCGATGCCATCGATGGTCGCCCCATGTCCATCGCTTCCGAGAAT
CAATACCCACGCCTCTACCATCCGGGCACTCCCCTGCCGATCTTCCCGCATGATCTTGACTTGGGTAAATCCATTTG
GTGGGGCGCCGGTTCCCACAACACCGCCGAACAAGCAGCACGCGATGGCGTTAACTTGATGAGCTCCACCCTCGTCG
CCGAAGCCACCGGCCAATCCTTCGGGGATCTGCAAGCCGATCAAATCGCGTTCTACCGCCAAGCTTGGAAAGAAGCC
GGACACGATTGGACCCCACGTGTGTCTGTCTCCAGGTCCATCTTTCCGATCGTCACCGACCGCGACCGTGAGCTTTT
CGGACTTCAGGGACAAGGCGGTGACCAAGTAGGAATCCTGGATGATACCCGATCCACGTTCGGTCGCAGCTACGCCG
GAAGTCCCGATGAACTCATCGACCAGCTCCAAGGAAGACAAAGCTGTGATGGAAGCCGACACCTTGATGCTCACCGC
CCCCAACCAAATGGGTGT
>RXA02531-nucleotide sequence C: downstream
TGAGATCAACGCGTCGATCCTGA

>>RXA01340-amino acid sequence
(1-990, translated) 330 residues
VGPVLAITPW NFPIAMATRK IAPALAAGCP VLVKPASETP LTMVKVGEII ASVFDTFNIP QGLVSIITTT
RDAELSAELM ADPRLAKVTF TGSTNVGRIL VRQSADRLLR TSMELGGNAA FVIDEAADLD EAVSGAIAAK
LRNAGQVCIA ANRFLVHESR AAEFTSKLAT AMQNTPIGPV ISARQRDRIA ALVDEAITDG ARLIIGGEVP
DGSGFFYPAT ILADVPAQSR IVHEEIFGPV ATIATFTDLA EGVAQANSTE FGLAAYGFSN NVKATQYMAE
HLEAGMVGIN RGAISDPAAP FGGIGQSGFG REGGTEGIEE YLSVRYLALP
>RXA01340-nucleotide sequence A: upstream
CAGAATACTTCCGTTGGTTTGCGGAAGAAGCAGTGCGCCTGCCCGGCCGCTACGGACAGTCACCTTCGGAATCGGTC
ACATCGCCGTCACCCGCGCACCC
>RXA01340-nucleotide sequence B: coding region
GTGGGACCAGTGCTGGCGATCACCCCATGGAATTTCCCCATCGCCATGGCCACCCGCAAAATCGCCCCAGCCCTGGC
CGCTGGTTGCCCCGTGTTGGTGAAACCTGCTTCCGAAACCCCACTGACCATGGTCAAAGTGGGGGAGATCATCGCCT
CCGTCTTTGATACCTTTAATATCCCGCAGGGCTTGGTCTCAATCATCACCACCACTCGAGATGCAGAGCTATCGGCA
GAACTCATGGCTGATCCTCGCTTGGCTAAAGTCACCTTCACTGGATCAACCAACGTGGGACGCATCCTGGTCCGCCA
ATCCGCGGACCGACTGCTGCGCACCTCCATGGAACTCGGCGGAAATGCAGCTTTTGTTATCGACGAAGCCGCAGACC
TCGACGAAGCCGTATCCGGTGCCATCGCCGCAAAACTCCGCAACGCCGGCCAAGTATGCATCGCAGCTAACCGTTTC
TTGGTTCATGAATCCCGCGCTGCCGAATTCACCTCAAAGCTGGCGACAGCCATGCAGAACACTCCCATTGGGCCGGT
GATTTCTGCCCGCCAACGCGACCGGATCGCAGCACTAGTGGATGAAGCCATCACCGACGGCGCCCGCCTCATCATCG
GTGGGGAGGTCCCCGACGGCTCCGGCTTCTTCTATCCAGCCACCATCTTGGCCGATGTCCCTGCACAGTCACGGATT
GTGCATGAGGAAATCTTCGGACCTGTGGCCACCATTGCCACTTTCACCGACTTGGCCGAAGGCGTTGCACAAGCAAA
TTCCACCGAATTCGGCCTCGCAGCCTACGGATTCAGCAACAATGTGAAAGCAACACAGTACATGGCGGAACACTTGG
AAGCCGGAATGGTCGGAATCAACAGAGGCGCCATCTCTGACCCAGCAGCACCTTTTGGCGGCATCGGACAATCCGGC
TTCGGCAGAGAAGGCGGAACCGAAGGAATCGAAGAATATCTCTCCGTGCGTTACCTCGCTTTGCCG
>RXA01340-nucleotide sequence C: downstream
TGACACATGAGCTGTCCGGTGAA

>>RXA01498-amino acid sequence
(1-1563, translated) 521 residues
MIKRLPLGPL PKELHQTLLD LTANAQDAAK VEVIAPFTGE TLGFGFDGDE QDVEHAFALS RAAQKKWVHT
TAVERKKIFL KVHDLVLKNR ELLMDIVQLE TGKNRASAAD EVLDVAITTR FYANNAGKFL NDKKRPGALP
IITKNTQQYV PKGVVGQITP WNYPLTLGVS DAVPALLAGN AVVAKPDLAT PFSCLIMVHL LIEAGLPRDL
MQVVTGPGDI VGGAIAAQCD FLMFTGSTAT GRILGRTMGE RLVGFSAELG GKNPLIVAKD ADLDKVEAEL
PQACFSNSGQ LCVSTERIYV EEDVYEEVIA RFSKAAKAMS IGAGFEWKYE MGSLINHAQL DRVSTFVDQA
KAAGATVLCG GKSRPDIGPF FYEPTVLADV PEGTPLLTEE VFGPVVFIEK VATLEEAVDK ANGTPYGLNA
SVFGSSETGN LVAGQLEAGG IGINDGYAAT WASVSTPLGG MKQSGLGHRH GAEGITKYAE IRNIAEQRWM
SMRGPAKMPR KVYSDTVATA LKLGKIFKVL P
>RXA01498-nucleotide sequence A: upstream
CAGTGGACAACTACTTGGCGGGTCTTAAATCAGCTGTGAAGGATTCTGCATAAGCTGGGCACCACACGAGCATCAGA
ACGCGAAACGAAGGTAAAAGCCC
>RXA01498-nucleotide sequence B: coding region
ATGATCAAACGTCTTCCTTTAGGTCCGCTGCCTAAAGAACTTCATCAGACTCTGCTTGATCTGACCGCAAATGCCCA
AGATGCGGCGAAAGTGGAGGTTATAGCGCCATTTACTGGCGAGACCCTCGGATTTGGTTTTGATGGTGATGAGCAAG
ACGTCGAGCATGCTTTTGCACTTTCAAGGGCAGCCCAGAAAAAGTGGGTGCACACCACGGCAGTGGAACGGAAGAAG
ATCTTCCTGAAGGTTCATGATCTGGTATTGAAAAACCGTGAGCTGCTCATGGACATCGTGCAGTTGGAAACAGGCAA
AAATCGAGCATCGGCTGCCGATGAGGTGTTGGACGTTGCGATCACCACCCGCTTCTACGCAAACAATGCAGGAAAGT
TTTTAAATGACAAGAAACGCCCCGGCGCGCTTCCGATCATCACGAAAAACACACAACAGTATGTGCCCAAGGGAGTG
GTCGGGCAGATCACGCCGTGGAATTACCCTTTAACTTTGGGAGTATCTGATGCTGTTCCGGCGCTGCTGGCAGGAAA
CGCAGTGGTGGCTAAACCTGACCTCGCGACACCTTTCTCCTGCTTGATCATGGTGCACCTGCTCATTGAAGCCGGTC
TGCCGCGTGATTTGATGCAGGTTGTCACCGGCCCTGGCGATATTGTTGGCGGTGCGATTGCAGCTCAGTGTGATTTC
CTCATGTTCACTGGATCCACGGCCACGGGCCGGATCTTGGGTCGGACAATGGGTGAGCGTTTGGTGGGTTTCTCTGC
GGAATTAGGCGGAAAGAACCCTCTTATTGTGGCCAAGGATGCAGATCTGGACAAGGTGGAAGCTGAGCTTCCGCAGG
CGTGTTTTTCCAACTCGGGGCAATTGTGTGTCTCCACTGAACGTATTTATGTCGAGGAAGACGTGTACGAGGAGGTG
ATTGCACGGTTTAGCAAGGCGGCGAAAGCCATGTCCATTGGTGCCGGATTTGAGTGGAAATATGAGATGGGTTCGTT
GATCAATCACGCGCAGCTGGATCGGGTGAGCACCTTTGTTGATCAGGCTAAAGCTGCGGGCGCCACGGTGCTGTGCG
GTGGCAAGTCACGCCCTGATATTGGTCCCTTCTTCTATGAGCCCACGGTATTGGCGGATGTCCCAGAGGGCACCCCA
CTGCTCACGGAGGAAGTCTTCGGGCCGGTGGTGTTCATCGAAAAGGTAGCCACACTGGAAGAAGCCGTCGATAAGGC
AAATGGCACGCCCTACGGCCTGAATGCGTCCGTCTTTGGGTCGTCGGAAACCGGCAATCTTGTTGCAGGCCAGCTGG
AAGCTGGCGGTATCGGTATTAATGATGGCTACGCCGCGACGTGGGCGAGCGTGTCCACGCCTCTGGGTGGCATGAAG
CAGTCGGGGCTGGGGCACCGCCATGGTGCGGAGGGAATTACAAAATATGCGGAGATCCGAAACATCGCGGAGCAGCG
CTGGATGTCTATGCGTGGGCCGGCCAAAATGCCGCGAAAGGTGTACTCAGACACCGTGGCCACAGCGCTAAAGCTGG
GCAAAATCTTTAAAGTTTTGCCG
>RXA01498-nucleotide sequence C: downstream
TAGCAAAAAGCCGGACCCTTGCT

>>RXA01182-amino acid sequence
(1-393, translated) 131 residues
VFNLFGRKTP RSNLRPPRGP GDTVRPEDLK FLMQWVQDKP FVEAFVEPET LVNEMSVVLV DAHGVFVRRR
IGGPKGIDVI AKKLGVPVYD VEETGYPQRM RERIEYERIL RKREEQKARR AKFERGENPD L
>RXA01182-nucleotide sequence A: upstream
GTTAAAACGGAAACTAATACCCCAAAGGATACCGATTCAATTTGTGATGTGTGGTGTTCGGGTCATATCAAGCTAAA
CAGATGCCCCCTACAATAGGCTT
>RXA01182-nucleotide sequence B: coding region
GTGTTCAATTTATTTGGTCGTAAAACTCCTCGCTCTAACCTCCGCCCACCACGCGGTCCGGGCGATACTGTGCGCCC
GGAAGATTTAAAATTCTTGATGCAATGGGTGCAGGATAAGCCATTTGTTGAGGCATTCGTTGAACCGGAAACGCTGG
TCAATGAGATGTCTGTCGTTTTGGTTGATGCTCATGGGGTTTTTGTCCGCCGAAGGATCGGCGGTCCCAAAGGGATT
GATGTTATCGCGAAAAAGCTCGGCGTTCCGGTTTATGATGTTGAGGAGACCGGTTACCCCCAAAGGATGCGCGAACG
CATTGAATATGAGCGCATCTTAAGAAAGCGTGAGGAACAAAAAGCTCGCCGCGCTAAATTTGAGCGCGGCGAGAATC
CTGATCTT
>RXA01182-nucleotide sequence C: downstream
TAACTAGCGTTTAGCTTTCCGAC

>>RXA00689-amino acid sequence
(1-627, translated) 209 residues
MRNTGQTCYI STRIIAPSSR YAEVVQTVAS TIAAGRQGDP YDEETVFGPV ASASQYSTVM SYIDSAREEG
ARVVAGGTRS ISLSEGLESG EFIQPTVFAD VTPDMRISRE EIFGPVISIL KYDDTNGVSE AIALANNTKF
GLGGLVFGAD EEQALEVARQ VDSGSVGINF FGSNHSAPFG GRHESGMGVE YGIEGLSAYL TYKSIHRTI
>RXA00689-nucleotide sequence A: upstream
ACTCCGACCAGTGACTTTAGAGCTAGGCGGAAAATCTTCCGCGATTATCCTTCCTGATGCAGACATGTCAGTACTCT
CGACGCGGTTGATTCGATCCTGT
>RXA00689-nucleotide sequence B: coding region
ATGCGCAACACTGGACAAACCTGCTACATCAGTACCCGGATTATTGCCCCTAGCTCACGCTATGCGGAAGTCGTACA
AACAGTGGCAAGCACTATCGCTGCAGGTAGACAAGGTGACCCCTATGATGAAGAAACGGTTTTTGGCCAGTTGCCA
GCGCCTCTCAGTACTCAACCGTCATGTCTTACATTGACTCCGCACGAGAGGAAGGTGCACGAGTGGTTGCAGGTGGA
ACCCGGTCAATCAGCCTTTCTGAAGGTTTAGAATCAGGCGAGTTTATCCAACCAACCGTGTTTGCCGATGTCACCCC
CGACATGCGGATATCACGCGAAGAAATCTTCGGCCCTGTTATTTCCATCCTAAAGTACGACGATACAAACGGTGTTT
CCGAAGCAATCGCACTAGCCAACAACACGAAATTCGGTCTCGGTGGCTTGGTATTTGGTGCGGATGAGGAACAAGCA
CTAGAAGTCGCCCGTCAAGTGGATTCTGGTTCCGTAGGCATCAACTTCTTCGGTTCCAACCATTCCGCCCCATTTGG
AGGACGCCACGAATCCGGTATGGGAGTGGAATACGGCATCGAAGGCCTCAGTGCTTACCTGACATACAAGAGTATTC
ACCGAACCATT
>RXA00689-nucleotide sequence C: downstream
TAGTTACTGAAAGTTCTCAGCTA

>>RXA00231-amino acid sequence
(1-789, translated) 263 residues
DRLRKVSFTG STPVGQQLLK KAADKVLRTS MELGGNAPFI VFEDADLDLA IEGAMGAKMR NIGEACTAAN
RFLVHESVAD EFGRRFAARL EEQVLGNGLD EGVTVGPLVE EKARDSVASL VDAAVAEGAT VLTGGKAGTG
AGYFYEPTVL TGVSTDAAIL NEEIFGPVAP IVTFQTEEEA LRLANSTEYG LASYVFTQDT SRIFRVSDGL
EFGLVGVNSG VISNAAAPFG GVKQSGMGRE GGLEGIEEYT SVQYIGIRDP YAG
>RXA00231-nucleotide sequence B: coding region
GATCGCCTTCGTAAAGTCTCCTTCACCGGCTCCACCCCAGTTGGCCAGCAGCTGCTCAAAAAGGCTGCCGATAAAGT
TCTGCGCACCTCCATGGAACTTGGTGGCAACGCACCTTTCATTGTCTTCGAGGACGCCGACCTAGATCTCGCGATCG
AAGGTGCCATGGGTGCCAAAATGCGCAACATCGGCGAAGCTTGCACCGCAGCCAACCGTTTCTTAGTCCACGAATCC
GTCGCCGATGAATTCGGCCGTCGCTTCGCTGCCCGCCTTGAAGAGCAAGTCCTAGGCAACGGCCTCGACGAAGGCGT
CACCGTGGGCCCCCTGGTTGAGGAAAAAGCACGAGACAGCGTTGCATCGCTTGTCGACGCCGCCGTCGCCGAAGGTG
CCACCGTCCTCACCGGCGGCAAGGCCGGCACAGGTGCAGGCTACTTCTACGAACCAACGGTGCTCACGGGAGTTTCA
ACAGATGCGGCTATCCTGAACGAAGAGATCTTCGGTCCCGTCGCACCGATCGTCACCTTCCAAACCGAGGAAGAAGC
CCTGCGTCTAGCCAACTCCACCGAATACGGACTGGCCTCCTATGTGTTCACCCAGGACACCTCACGTATTTTCCGCG
TCTCCGATGGTCTCGAGTTCGGCCTAGTGGGCGTCAATTCCGGTGTCATCTCTAACGCTGCTGCACCTTTTGGTGGC
GTAAAACAATCCGGAATGGGCCGCGAAGGTGGTCTCGAAGGAATCGAGGAGTACACCTCCGTGCAGTACATCGGTAT
CCGGGATCCTTACGCCGGC
>RXA00231-nucleotide sequence C: downstream
TAGCATCTGCCCCTTTACAAATC

>>RXA00659-amino acid sequence
(1-621, translated) 207 residues
VEASDGGTFD VENPATGETI ATLASATSED ALAALDAACA VQAEWARMPA RERSNILRRG FELVAERAEE
FATLMTLEMG KPLAEARGEV TYGNEFLRWF SEEAVRLYGR YGTTPEGNLR MLTALKPVGP CLLITPWNFP
LAMATRKVAP AIAAGCVMVL KPARLTPLTS QYFAQTMLDA GLPAGVLNVV SGASASAISN PIMEDDR
>RXA00659-nucleotide sequence A: upstream
GCTGGATACGAAAAGTGAAGGAAAATAACGCATCATGACTATTAATGTTTTCGAACTACTTGTCAAAAGTCCCACGG
GTCTACTGATTGGTGATTCCTGG
>RXA00659-nucleotide sequence B: coding region
GTGGAAGCATCCGACGGCGGTACTTTCGATGTGGAAAACCCAGCGACGGGTGAAACAATCGCAACGCTCGCGTCTGC
TACTTCCGAGGATGCACTGGCTGCTCTTGATGCTGCATGCGCTGTTCAGGCCGAGTGGGCTAGGATGCCAGCGCGCG
AGCGTTCTAATATTTTACGCCGCGGTTTTGAGCTCGTAGCAGAACGTGCAGAAGAGTTCGCCACCCTCATGACCTTG
GAAATGGGCAAGCCTTTGGCTGAAGCTCGCGGCGAAGTCACCTACGGCAACGAATTCCTGCGCTGGTTCTCTGAGGA
AGCAGTTCGTCTGTATGGCCGTTACGGAACCACACCAGAAGGCAACTTGCGGATGCTGACCGCCCTCAAGCCAGTTG
GCCCGTGCCTCCTGATCACCCCATGGAACTTCCCACTAGCAATGGCTACCCGCAAGGTCGCACCTGCGATCGCTGCA
GGTTGTGTCATGGTGCTCAAGCCAGCTCGACTTACCCCGCTGACCTCCCAGTATTTTGCTCAGACCATGCTTGATGC
CGGTCTTCCAGCAGGTGTCCTCAATGTGGTCTCCGGTGCTTCCGCCTCTGCGATTTCCAACCCGATTATGGAAGACG
ATCGC

>>RXA02192-amino acid sequence
(1-522, translated) 174 residues
IPDIGFGVFQ TPPDETRNSV NAALEAGYRH IDTAAAYGNE REVGEAIAAS GIGRDEITIE TKIWVTDYGF
EETLHAFDKA TGKLGVDTLD ILILHQAVPS SFDRTIAAYK ALEKLLFDGA VRAIGVSNFM PEHLDKLLLE
TSIVPALNQI ECHPYFQQRD VLARNEQLGI LTQA
>RXA02192-nucleotide sequence B: coding region
ATTCCCGACATTGGATTTGGTGTCTTCCAAACCCCACCCGATGAAACCCGAAACTCCGTTAACGCTGCTCTTGAAGC
CGGCTATCGCCACATCGACACCGCGGCCGCATACGGCAATGAACGTGAAGTCGGTGAAGCAATCGCAGCATCCGGCA
TTGGCCGCGACGAGATCACCATCGAAACCAAAATCTGGGTGACCGACTACGGCTTCGAGGAAACTCTCCACGCATTC
GACAAGGCCACAGGCAAGCTTGGTGTCGATACACTGGACATTTTGATCTTGCACCAGGCAGTGCCAAGCAGCTTTGA
TCGCACCATCGCCGCCTACAAGGCGCTAGAGAAGCTGCTTTTCGACGGCGCGGTGCGGGCAATCGGAGTCAGTAATT
TCATGCCAGAGCACCTGGACAAACTCCTTTTGGAAACCTCCATTGTCCCAGCTCTGAACCAAATCGAATGCCACCCC
TACTTCCAGCAGCGTGACGTGCTTGCCCGCAATGAGCAGCTTGGCATTTTGACTCAGGCG

Claims

1. An isolated nucleic acid molecule from *Corynebacterium glutamicum* encoding an HA protein, or a portion thereof.

2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule encodes an HA protein involved in the production of a fine chemical.

3. An isolated *Corynebacterium glutamicum* nucleic acid molecule selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

4. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth in Appendix B.

5. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth in Appendix B.

6. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

7. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A.

8. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1-7 under stringent conditions.

9. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-8 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.

10. A vector comprising the nucleic acid molecule of any one of claims 1-9.

11. The vector of claim 10, which is an expression vector.

12. A host cell transfected with the expression vector of claim 11.

13. The host cell of claim 12, wherein said cell is a microorganism.

14. The host cell of claim 13, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

15. The host cell of claim 12, wherein the expression of said nucleic acid molecule results in the modulation in production of a fine chemical from said cell.

14-07-99

16. The host cell of claim 15, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.

17. A method of producing a polypeptide comprising culturing the host cell of claim 12 in an appropriate culture medium to, thereby, produce the polypeptide.

18. An isolated HA polypeptide from *Corynebacterium glutamicum*, or a portion thereof.

19. The polypeptide of claim 18, wherein said polypeptide is involved in the production of a fine chemical production.

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, or a portion thereof.

22. The isolated polypeptide of any of claims 18-21, further comprising heterologous amino acid sequences.

23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth in Appendix A

24. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

25. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 12 such that the fine chemical is produced.

26. The method of claim 25, wherein said method further comprises the step of recovering the fine chemical from said culture.

27. The method of claim 25, wherein said method further comprises the step of transfecting said cell with the vector of claim 11 to result in a cell containing said vector.

28. The method of claim 25, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

29. The method of claim 25, wherein said cell is selected from the group consisting of: *Corynebacterium glutamicum, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum,*

Corynebacterium acetophilum, Corynebacterium ammoniagenes, Corynebacterium fujiokense, Corynebacterium nitrilophilus, Brevibacterium ammoniagenes, Brevibacterium butanicum, Brevibacterium divaricatum, Brevibacterium flavum, Brevibacterium healii, Brevibacterium ketoglutamicum, Brevibacterium ketosoreductum, Brevibacterium lactofermentum, Brevibacterium linens, Brevibacterium paraffinolyticum, and those strains set forth in Table 3

30. The method of claim 25, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.

31. The method of claim 25, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.

32. The method of claim 25, wherein said fine chemical is an amino acid.

33. The method of claim 32, wherein said amino acid is drawn from the group consisting of: lysine, glutamate, glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.

34. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-9.